MATHEMATICA Policy Research

REPORT

FINAL

Evaluation of the Round Two Health Care Innovation Awards (HCIA R2): Third Annual Report

June 2018

Boyd Gilman	Laurie Felland	Cyrus Jadun
Nancy McCall	Rachel Kogan	Natalie Graves
Karen Bogen	Peggy O'Brien-Strain	Theresa Feeley-Summerl
Arnold Chen	Rumin Sarwar	Lori Timmins
Leslie Conwell	Robert Schmitz	William Mulhern
Stacy Dale	Joe Zickafoose	Randall Brown

Submitted to:

Centers for Medicare & Medicaid Services
Innovation Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Project Officer: Jean M. Gaines, COR and Evaluation Co-leader for Specialty Care and Older Populations Group; Patricia Markovich, Evaluation Co-leader for Models of Community-Based Care Group
Contract Number: CMMI-500-2014-0000341

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown

Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

EXECUTIVE SUMMARY

Introduction

Under contract to the Centers for Medicare & Medicaid Services (CMS), Mathematica Policy Research is evaluating the 38 unique programs funded under Round Two of the Health Care Innovation Awards (HCIA R2). Thirty-nine organizations received three-year cooperative agreements, beginning in September 2014, to implement their proposed models for improving the quality of care and health, and for lowering the cost of care for Medicare, Medicaid, and Children's Health Insurance Program beneficiaries. However, in September 2016, one awardee terminated its cooperative agreement with the Center for Medicare & Medicaid Innovation (CMMI) and withdrew from the HCIA R2 initiative, so this report includes only 38 programs. Mathematica is evaluating the extent to which the remaining awardees successfully implemented their programs and are accomplishing their goals. The programs vary widely in the interventions provided, the populations served, the types of organizations involved, and the number of individuals enrolled.

The findings presented in this report—the third of four evaluation reports of the HCIA R2 cooperative agreements—cover four evaluation components: (1) implementation effectiveness, (2) developing payment models, (3) planning for sustainability, and (4) preliminary work on evaluating program impacts. The findings for the first three components reflect all 38 awardees and are based on the awardees' self-reported qualitative and survey information collected through the end of the three-year cooperative agreements (August 31, 2017). The fourth component included an assessment of the evaluability of each awardee's program, the initial identification of comparison groups for three awardees, and an interim evaluation of impacts for three awardees—two that implemented randomized controlled trials and one for which we identified a credible comparison group in time to complete the analyses for this report. The impact analyses were limited to first-year outcomes for beneficiaries who enrolled in an awardee's program during the first 27 months after the award of the HCIA R2 cooperative agreements.

In summer 2017, CMS awarded no-cost extensions to 30 awardees, ranging from three months to one year. The final report will present changes or progress made during the extension period, along with final impact estimates for all awardees for which we can conduct a credible evaluation.

In short, we find that about half of the awardees met at least 90 percent of their enrollment target and almost two-thirds implemented their service delivery models effectively, awardees' progress sustaining their programs and developing the payment models to support them was mixed, and the limited early impact analyses found little evidence of favorable effects on utilization or expenditures. The following pages summarize the key findings provided in this report in greater detail.

Implementation effectiveness

- About half of the awardees (20) met at least 90 percent of their final enrollment projections (set at the end of Year 2). Another one-third (13) met 65 to 90 percent of their final projections. Five awardees failed to meet 65 percent of their final enrollment targets.
  - Awardees could, and most did, revise their enrollment goals during the cooperative agreement. The 24 awardees that lowered their initial goals were more likely to meet their final enrollment projections than were the 14 that either increased their projections or left them unchanged during the cooperative agreement.
  - When measured against the enrollment goals set by awardees at the end of Year 1, enrollment performance was weaker. Only 14 awardees met at least 90 percent of their end-of-Year-1 goals, 16 met 65 to 90 percent, and 8 failed to meet 65 percent of their goals.
- We concluded that 24 of the awardees implemented their service delivery models effectively. That is, any problems that the awardee might have encountered with enrolling participants, delivering services, employing and training staff, and engaging providers and patients were not deemed serious enough to threaten the awardee's ability to improve its targeted outcomes. The other 14 awardees were partly effective in implementing their service delivery models, implying that shortcomings in one or more of these areas might have impeded their ability to improve outcomes.
- Programs with these structural features were more likely to have been implemented effectively: (1) targeted chronic rather than acute diagnoses, (2) focused on changing providers' rather than patients' behaviors, (3) served single or local areas rather than large or multiple areas, and (4) had piloted-tested their model rather than implemented an untried program.
- The evaluation team identified six activities or situations associated with effective service delivery:
  - **Creating a supportive internal environment** through a team-oriented culture, strong communication and supportive structures, and the presence of program champions
  - **Obtaining the active participation of providers and community organizations** by understanding their constraints and needs, and by offering services perceived as valuable
  - **Keeping participants engaged in the program** by building trust and rapport with them through personal hand-offs between clinicians and staff and open communication via texting, tailoring services to participants' needs, and furnishing services perceived as valuable
  - **Attending to ongoing staffing issues** by hiring, training, and retaining enough staff with the appropriate skills for the program
  - **Deploying health information technology** (health IT) to facilitate the use of clinical data in patients' care, support the flow of patients' data among providers and between providers and patients, and offer patients ways to engage with and manage their own care

- **Having external circumstances favorable to the program,** such as regulations and policies aligned with program goals and an adequate supply of health care resources

Developing payment models and planning for sustainability

CMS expected awardees not only to test their interventions but also to build support for them among potential payers so that the models can continue to provide innovative services after the cooperative agreements end. Part of this effort requires developing feasible arrangements for these insurers and managed care organizations to pay for the program's services.

- Seven awardees appeared to have the necessary plans and resources in place to continue their programs largely intact after the cooperative agreement ended. Twenty-five awardees expected to sustain at least parts of their programs but were at different stages of acquiring the funding necessary to do so. Sustainability was uncertain or unlikely for the remaining six awardees.
- Six factors associated with greater progress in sustainability planning include the following:
  - **Having some evidence demonstrating the program's likely value** to payers
  - **Engaging both the internal leaders and frontline staff** needed to continue and potentially expand their programs
  - **Obtaining support from payers or other funders**, including state legislators and agencies (particularly Medicaid), public and private insurers, and private foundations
  - **Having a program design that is adaptable or flexible** enough to accommodate changes or challenges faced by implementing organizations
  - **Having program features and objectives that match the mission of the implementing organization**
  - **Having a program that is aligned with federal and state policies,** especially value-based payment approaches
- We found no clear association between the awardees' plans to sustain, scale, or replicate their programs and overall implementation effectiveness. However, awardees that were effective in staff recruitment and training were more likely than other awardees to have advanced sustainability plans.
- By the end of the third year, 37 awardees had developed 46 different payment models for their programs. Of the 46 models, 36 incorporated one or more aspects of an alternative payment model, moving away from fee-for-service and toward value-based purchasing. Eight awardees have executed a contract with at least one payer, and an additional 15 were in negotiations with payers.
- Factors associated with greater progress in developing payment models include (1) consulting with experts in payment strategies, (2) leveraging existing alternative payment model initiatives, (3) collaborating with payers, (4) learning from peer groups, and (5) acquiring data to calculate costs and demonstrate value.

- Having a payment model contract in place was important for sustainability, but it was not always necessary or sufficient. Only three of the eight awardees with a payment model contract in place seemed ready and able to sustain their whole programs at most implementing sites. The other five awardees planned to sustain their programs as well, but only for some components of their interventions or at only some of their implementing sites. Four awardees without a payment contract in place also seemed certain of their ability to sustain their programs by using other sources of funding.
- We did not detect a meaningful relationship between implementation effectiveness and awardees' progress in developing their payment models. This general lack of association is less surprising than might be expected, given that executing payment models depends more on promotional skills and negotiations among high-level executives, whereas implementation effectiveness depends on the ability of frontline staff to provide care.

Impact evaluation

- We expect to produce rigorous beneficiary-level impact estimates for 23 awardees (9 for Medicare only, 8 for Medicaid only, and 6 for both Medicare and Medicaid). Well-matched comparison groups have been selected for 3 awardees to date, and 2 others have control groups arising from a randomized controlled trial.
- The early impact findings for 3 of the 23 awardees (shown in Table ES.1) are mixed and should be interpreted with caution because (1) the analysis included only beneficiaries enrolled in a program during roughly the first two years of the cooperative agreement and (2) outcomes were measured during only the first year after enrollment. Estimated impacts on total expenditures were not statistically significant for any of the 3 awardees. For 1 of the 3 awardees (the University of Illinois at Chicago, or UIC), statistically significant estimates of program impacts were favorable for several outcomes. UIC was not among those awardees that we assessed as having effectively implemented its program.
- We will not be able to produce impact estimates for the remaining 15 awardees because they have too few enrollees, their eligibility criteria cannot be replicated for a comparison group, and/or their primary outcomes are not obtainable from claims data. For these awardees, we will conduct alternative analyses, such as a comparison of outcomes among treatment beneficiaries in the pre- and post-award periods.

Table ES.1. Summary of percentage effect of three awardees' programs on expenditures, rates of utilization, and likelihood of any acute care utilization during a 12-month follow-up period

Expenditures and utilization measures	UCSF (247/132)	UIC[a] (3,131/3,128)	NYC H+H (2,042/6,145)
Comparison group	**RCT**	**RCT**	**Matched group**
Medicare or Medicaid expenditures			
Total	-8% N/S	-8% N/S	3% N/S
Acute inpatient	N/S	N/S	N/S
Inpatient other			N/S
Outpatient			N/S
Physician services			-10%***
Home health			N/S
SNF			N/S
Hospice			N/S
Durable medical equipment			N/S
Outpatient ED		N/S	
Prescription drugs		N/S	
All other Medicaid-covered services		N/S	
Service utilization rates			
Acute hospital admissions	N/S	N/S	N/S
Outpatient ED visits	N/S	-7%*	N/S
Primary care visits in ambulatory settings			-13%***
Specialist visits in any setting			-10%***
Specialist visits in ambulatory settings			-11%***
Percentage of beneficiaries with a service use			
Percentage with a hospitalization	N/S	N/S	N/S
Percentage with an outpatient ED visit or observation stay	N/S	-2.4*	+2.2*
Percentage with a readmission among all beneficiaries	N/S		N/S

Note: Percentage effect = Impact regression coefficient/(Predicted treatment group mean – impact regression coefficient)*100. N/S = estimate not statistically significant at .10 level; shaded cells = not evaluated.

Treatment and comparison group sample sizes (n_T/n_C) appear in parentheses under the awardees' names.

[a] For UIC, the outcomes are Medicaid expenditures per beneficiary per month.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department; NYC H+H – New York City Health and Hospitals; RCT = randomized controlled trial; SNF = skilled nursing facility; UCSF = University of California, San Francisco; UIC = University of Illinois at Chicago.

Next steps

The primary focus of the evaluation over the next 18 months will shift from assessing implementation to measuring impacts. Specifically, we will do the following:

1. **Update our implementation findings**—including the awardees' payment models and broader strategies for sustaining, scaling, and replicating their programs—based on a review of the awardees' final documents and follow-up interviews with awardees with a no-cost extension.

2. **Assess the participants' experiences and satisfaction with the program** as well as their perceptions of the effect of the service delivery model on access to care, quality of care, and health outcomes, based on data from the patient survey fielded in summer 2017.
3. **Continue our impact analyses** by (1) developing matched comparison groups for the entire set of eligible beneficiaries who ever enrolled in a program for which an impact analysis is feasible, (2) estimating impacts on Medicare (and Medicaid) beneficiaries for these awardees by using conventional analyses and Bayesian models, and (3) performing sensitivity and robustness checks on the results.
4. **Identify and implement strategies for evaluating programs for which we will not be able to conduct a rigorous impact analysis** using awardee-provided and claims data for participants, if possible.

The final evaluation report that is due to CMMI toward the end of 2019 will present results of the Years 4 and 5 implementation, survey, and impact activities.

I. INTRODUCTION

A. Background and purpose of the HCIA R2 awards

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of Health Care Innovation Awards (HCIA R2) to 39 organizations. The selected awardees each proposed innovative ways to improve the quality and lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries. The awardees' goals were to (1) reduce Medicare, Medicaid, or CHIP costs in outpatient or post-acute settings; (2) improve care for patients with special needs; (3) test new financial and clinical models for specific provider types; and (4) improve the health of specific populations by enhancing patients' engagement and improving disease prevention, wellness, and comprehensive care. In contrast to the first round of HCIA, CMS also sought new payment models to support the service delivery models funded by this initiative. In February 2015 (five months after award of the HCIA R2 cooperative agreements), CMMI selected Mathematica Policy Research and its partners to evaluate the HCIA R2 programs.

Program operations will not end until August 2018 for some awardees. The initial three-year cooperative agreements ended in August 2017. However, in June 2017, CMS awarded no-cost extensions to 30 awardees, ranging from three months to one year. Most awardees continue to enroll and serve beneficiaries during this period. Others are using their extensions to complete administrative and local monitoring and evaluation activities. One awardee voluntarily terminated its cooperative agreement with CMMI and withdrew from the HCIA R2 initiative at the end of the second year and is not included in this report.

B. Evaluation goals and purpose of this report

The goals of the five-year evaluation are to assess whether and how the programs are transforming the delivery and financing of health care services, as well as how these changes are improving the coordination, cost, and quality of care. The evaluation has six objectives to help CMMI achieve these goals:

1. **Describe the implementation experience of each awardee** and assess the barriers and facilitators associated with the awardee's success in promoting change.
2. **Assess for each awardee the experience of participants, the attitudes of clinical and nonclinical staff** toward the model and their work, and participant and staff perceptions of the intervention's effects on the processes and outcomes of care.
3. **Assess the effects of each model on health care costs, utilization, and quality of care** by using the same methodologies and outcome measures across awardees when possible, plus additional outcome measures tailored to each program as appropriate.
4. **Synthesize the findings from the implementation and impact evaluations** of each program with input from key stakeholders to (1) identify what model components appear to be most critical to success and how the administrative, geographic, and organizational contexts influenced this success; and (2) provide evidence to CMMI about the sustainability, scalability, and replicability of each type of model.

5. **Describe the awardees' payment model designs** and their experience in developing and testing the models, including the challenges they faced and the strategies they used to address them.
6. **Conduct a meta-evaluation of the awardee-specific results,** searching for program features consistently associated with both successful implementation and program impacts within subgroups of similar awardees and possibly across the entire portfolio of awardees.

This is the third of four annual evaluation reports we will submit to CMMI. The first two reports addressed the first and fifth evaluation objectives (describe the implementation of the service delivery models and the development of the payment models) and laid the foundation for addressing the other objectives in the future reports. The two reports are publicly available on CMS's website.[1]

This report draws on the two previous reports and on new data and analyses to accomplish three general purposes:

1. **Summarize the findings on the awardees' success in implementing the service delivery models,** identify the barriers and facilitators associated with implementation success, and highlight the implications of the findings for the impact analyses.
2. **Summarize the findings to date on the impact evaluation** by updating our assessment of evaluability, describing the comparison groups selected for three awardees, and presenting interim impact findings for three awardees (two of which implemented a randomized controlled trial).
3. **Synthesize the awardees' progress toward sustaining their programs** after the end of the cooperative agreement, based on the development of their payment models and the steps taken to develop and implement their sustainability plans.

We based our analysis of the implementation experience and effectiveness on a review of the awardees' self-reports and telephone interviews with program administrators and frontline staff. We reviewed the self-reported awardee documents from all 12 program quarters (September 2014 to August 2017). We conducted annual interviews with staff from up to three implementing sites per awardee toward the end of each of the three years of the cooperative agreement. In addition, we used program enrollment data (through August 2017) that the awardees provided to the implementation and monitoring contractor to describe how the number of participants compared with the awardees' final cumulative enrollment projections. We also relied on responses to surveys of non-clinician and clinician staff who participated in the programs. We fielded the non-clinician staff survey from July to October 2016 and the clinician survey from March to June 2017. Our assessment of implementation effectiveness is based on self-reported information; we were unable to independently verify the awardees' input.

Our preliminary impact analyses are based on Medicare (and, for one awardee, Medicaid) enrollment and claims data for participants who entered the program before November 30, 2016.

[1] The first and second evaluation reports can be accessed at https://downloads.cms.gov/files/cmmi/hcia2-yroneevalrpt.pdf and https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf, respectively.

The calculation of baseline characteristics and the analysis of comparison groups are based on finder files available when the analyses were conducted, which ranged from May through August 2017.

C. Unique challenges of evaluating the HCIA R2 awardees

Although fostering innovation from the ground up enables CMS to test a wide range of locally initiated models simultaneously, it also introduces a unique set of evaluation challenges. The first is the sheer diversity of programs. As discussed in Chapter II (and highlighted in our previous reports), the 38 programs in operation at the end of the three-year agreement varied widely in terms of service delivery models, target populations, and care settings. This diversity made it particularly challenging to synthesize findings across awardees and draw generalizable conclusions about the factors that led to implementation success or failure. To facilitate our cross-cutting analysis, we therefore grouped awardees by organizational and program features and looked for differences in the implementation experience based on these characteristics.

Second, unlike traditional demonstrations, CMS encouraged awardees to innovate and allowed them to make (sometimes major) program changes during the cooperative agreement. In some cases, the resulting absence of clear, standardized intervention protocols led to wide variation in how sites participating in the same program provided services. The emphasis on innovation and evolution rather than fidelity to an unchanging model sometimes made it difficult to define clearly the intervention being evaluated, and how and when the intervention changed. We attempted to document the major program changes and relevant history in the individual awardee narratives.

Third, conducting 38 separate evaluations limited the amount of time and resources we could commit to collecting and verifying data for any single evaluation. We did conduct annual site visit interviews with frontline staff at up to three purposively chosen sites per awardee. For some awardees, this represented a small number of their implementing sites. We corroborated some of the comments from the frontline staff by analyzing the responses to the clinician and nonclinician staff surveys. But beyond a review of the survey data and the awardees' quarterly progress reports, we had no way of verifying whether the information we collected was representative of the overall intervention staff.

Finally, the 38 evaluated programs differ widely in how credibly their impacts can be estimated and the amount of data available at this stage. These constraints limit the number of awardees for which we could include results in this report, the type of impact analyses that we could conduct, the number of enrollees included in the analyses, and the time periods covered.

D. Road map to the report

The remainder of this report presents the following:

- An updated overview of the 38 remaining HCIA R2 programs, highlighting the innovations in the service delivery models **(Chapter II)**

- Findings from our implementation evaluation in three areas: (1) assessment of implementation effectiveness, (2) factors associated with implementation effectiveness, and (3) implications of implementation effectiveness for the impact analyses **(Chapter III)**
- Findings from our impact evaluation to date, including (1) updated assessment of the evaluability of each awardee's program; (2) characteristics of enrollees based on an analysis of Medicare (and, for some awardees, Medicaid) enrollment and claims data for the year before enrollment, (3) an assessment for three awardees of the similarity of the selected comparison group to the treatment group on baseline characteristics, and (4) interim impact findings for three awardees **(Chapter IV)**
- An update to our evaluation of the development and implementation of payment models and sustainability plans **(Chapters V and VI, respectively)**
- Next steps in our implementation and impact evaluations, including an analysis of the beneficiary survey **(Chapter VII)**

Appendix A shows the number of respondents and response rate for each of the 21 awardees that participated the HCIA R2 patient survey. **Appendix B** describes the characteristics of the three matched comparison groups and two randomized control groups, and assesses the balance of each of the three treatment and comparison group pairs presented in this report. **Appendix C** provides technical details on our impact evaluation methodology. The 38 awardee-specific narratives, which formed the basis for the implementation and impact syntheses in the main body of this report, are included in **Appendix D**.

II. OVERVIEW OF HCIA R2 AWARDEES AND THEIR PROGRAMS

A. Introduction and summary of key findings

This chapter provides a brief overview of the 38 HCIA R2 awardee programs at the end of the third program year. The two previous annual reports described the programs in detail. As this annual report turns to findings on the factors associated with effectively implemented programs and their implications for the impact analyses, this overview highlights the range of the awardees' programs and focuses on three dimensions of all of them: (1) key innovations, (2) target population, and (3) expected impacts on core outcome measures. (The individual awardee narratives present additional detail on the characteristics of the awardees and their programs.)

In summary, we identified six high-level groups of key innovations, with the innovations differing by target population. There was less variation across the core outcome measures than there was across the target populations.

- More than half of the awardees offered innovative approaches to care coordination or care management. These innovations often targeted populations with high-risk chronic conditions or behavioral health and cognitive disorders. The awardees also extended the responsibility for providing care coordination services to providers less commonly linked to primary care or working outside of traditional medical practices.
- Awardees targeting adults who sought or who recently received acute or sub-acute care structured their innovations around care transitions or patient education and engagement.
- The expected effects on core outcomes were tied more to the target population than to the type of innovation. Awardees targeting high-risk chronic conditions were most consistently expected to reduce admissions and emergency department (ED) visits; those targeting primary and preventative care were least likely to affect these outcomes.

A. Key innovations

By definition, all 38 HCIA R2 awardee programs operating at the end of the third program year were designed to be innovative approaches to service delivery that improve care and lower costs. The second annual report described the major components of the service delivery models; Table II.1 summarizes the key innovative components for each awardee, based on the implementation evaluation. Table II.2 describes the key innovations. In many cases, the innovations built on existing models of service delivery and extended them to individuals with particularly challenging health conditions, to providers less commonly included in health care transformation, or to less traditional settings or underserved communities.

Among the most common building blocks for the awardees' innovations were care coordination and care management, patient education and engagement, and support for the transitions of care. Virtually all of the awardees with a patient-based intervention incorporated one or more of these components into their programs. Other common innovations, particularly for provider-based interventions, involved health information technology (IT), tele-health, and expanded roles for providers. As shown in Table II.1, these categories are not mutually

exclusive; many awardees incorporated multiple types of innovations. (Some awardees incorporated other components in their service model.)

More than half of the programs included innovative approaches to **care coordination and care management** (20 awardees). This includes two-thirds (8 of 12) of the programs that served children and youth and three programs that focused on behavioral health. In general, care coordination and care management programs organize care across providers and support multiple health care and support needs for patients and their caregivers. The innovations in this area involved (1) tailoring coordination approaches to chronic conditions that are particularly challenging or that involve specific care guidelines, (2) extending the programs to individual providers who are less commonly linked to primary care, or (3) providing care coordination outside of traditional medical practices. Examples of tailoring care coordination to specific conditions include programs that applied services to individuals with advanced chronic kidney disease, chronic obstructive pulmonary disorder (COPD), and hepatitis C (Northwell Health, Ventura County Health Care, and the Fund for Public Health in New York), and to children with complex medical conditions (Seattle Children's Hospital). Programs that involved new types of providers included those that extended care coordination efforts to better link primary care to dentistry (Altarum), palliative care (Four Seasons Compassion for Life), or pediatricians who treat children with complex medical needs (Boston Medical Center). Several awardees incorporated home visits into their care coordination models, including care for individuals with dementia (Johns Hopkins University and the University of California at San Francisco). Awardees that extended their programs beyond traditional settings included the Children's Home Society, which placed its intervention in community schools. New York City Health + Hospitals and Detroit Medical Center used a coordinated care model in or associated with a hospital ED. For youth transitioning out of foster care, Amerigroup offered care coordination as part of a larger package of wraparound services along with educational support and employment coaching.

Awardees implemented innovative models in six areas:

1. Care coordination/care management (20 awardees)
2. IT/decision support (11 awardees)
3. Care transition (10 awardees)
4. Patient engagement and education (10 awardees)
5. Extended provider roles (6 awardees)
6. Tele-health (5 awardees)

Patient engagement and education programs were another common focal point for the awardees' innovations. Patient education and engagement empowers patients and families to manage their (or a family member's) health conditions. Almost all (17 of 20) of the care coordination models incorporated patient engagement and education. A few awardees used patient education and engagement to support the optimal treatment of specific conditions. For example, the University of California at San Diego provided patient education and health coaching in connection with medication "bundles" to prevent heart attacks and strokes, and the University of Michigan provided pre-surgical patient education to promote better surgical outcomes. Several patient engagement programs incorporated innovative uses of technology for peer support and mentoring, including the VillageCare's program, which used social media and other technology to promote treatment adherence.

Several of the awardees' innovations redesigned **care transition programs.** Whereas care coordination programs generally provide ongoing care, care transition programs offer care coordination that supports patients transitioning from one care setting to another (such as from a hospital to home or to a skilled nursing facility [SNF]); the resulting handoffs from one provider to the next are part of the coordination package. Awardee innovations in this area include the CareChoice Cooperative's program that improved transitions from nursing facilities into the community, and the Icahn School of Medicine's program that addressed acute rehabilitation needs in the home. Also included in the care transition programs was the City of Mesa Fire and Medical Department. In addition to its medical and behavioral health program to prevent transport to the emergency department, the awardee also created mobile community medicine teams, which offered in-home support on issues such as medication adherence and fall prevention among low-acuity patients who were at high risk for hospital readmission.

VillageCare's use of a private social network and virtual video support groups was one example of many of the care coordination, patient engagement, and care transition program models that also incorporated innovative **health IT.** These programs commonly leveraged electronic medical records (EMRs) and collaborative care platforms to formalize referrals to specialists, communicate across providers, and track access to supportive services. In several cases, innovative health IT applications offered decision support for providers and facilitated shared decision making for patients and their families. Examples include the tools for evidence-based treatment of new-onset lower back pain, or LBP (the University of North Carolina at Chapel Hill) and the SMARTCare tools for ischemic heart disease (the American College of Cardiology Foundation).

Health IT was also an important ingredient in innovative **tele-health** approaches. Eight awardees incorporated a tele-health component into their programs, and it was a core innovation in five. Four of these awardees incorporated tele-health into innovative programs for rural patients and their providers. The technology enabled awardees to offer remote access to health care providers for SNF residents, to clinical follow-up after hospital discharge, to critical care specialists for emergency care, and to neurologists for individuals with neuro-emergent conditions. Tele-health innovations were not limited to rural areas. For example, in its CHECK program, the University of Illinois, Chicago, used tele-health technology for remote access to mental health services for Chicago area residents.

A final major area of innovation was **expanded roles for providers,** a key feature in four programs. As noted earlier, the Mesa Fire and Medical Department expanded the role of emergency responders to treating low-acuity care patients to prevent hospital readmissions. Yale University also expanded the function of the paramedic workforce to providing in-home interventions to prevent falls and to connect individuals with primary care. Other awardees expanded the role of primary care providers to include oral health care and the role of pharmacists to serving as extensions of those who manage care in patient-centered medical homes. Finally, although it included components of care management and care coordination and patient engagement, the key innovation in the National Association of Children's Hospitals and Related Institutions' program was its national learning collaborative to build capacity nationwide to serve children with complex medical needs.

B. Target population

As in our second annual evaluation report, we structured much of our data collection and analysis around the six groups of awardees listed below and shown in Table II.1. We assigned awardees to groups based on the age of their primary target populations and by the types and severity of their health conditions. These characteristics are often linked to types of care settings, providers, and in some cases, payers.

1. **Youth and young adults with chronic or complex medical conditions.** The five awardees in this group served children and youth exclusively, except for the University of Illinois, which served youth and adults younger than 25. All of the awardees included care management, care coordination, and patient or family engagement in their models, although the key innovations were split between care coordination and health IT for decision support, such as the National Association of Children's Hospital's learning collaborative. The target populations of these awardees faced complex or chronic medical conditions, sometimes combined with other challenges. These awardees typically served youth covered by Medicaid or CHIP.

Most awardees targeted one of 6 population groups:

1. Youth and young adults with chronic or complex medical conditions (5 awardees)
2. Adults with chronic physical conditions—high risk of costly acute care utilization (8 awardees)
3. Adults with chronic physical conditions—lower risk of costly acute care utilization (6 awardees)
4. People (adults, young adults, or children) with behavioral health or cognitive disorders (5 awardees)
5. Adults who sought or who recently received acute or sub-acute care (6 awardees)
6. People (adults, young adults, or children) who need primary or preventive care (8 awardees)

2. **Adults with chronic physical conditions—high risk of costly acute care utilization.** Eight awardees targeted primarily adults with a high risk of costly acute care utilization. They fall into two general groups: (1) those with late-stage chronic conditions, including chronic kidney disease, cancer, or life-limiting conditions needing palliative care (Four Seasons, Northwell, and University Hospitals Cleveland Medical Center); and (2) those with at least one chronic condition and at least one ED visit, hospitalization, or hospital readmission during the year before enrolling in an HCIA R2 program (Detroit Medical Center, New York City Health + Hospitals, the University of California at San Diego, and the University of Kansas Hospital Authority). Almost all of these programs provided care management, but the awardees implemented it in innovative ways and combined it with other service delivery components, such as patient engagement, patient education, tele-health, or medication therapy management. A variety of payers, including Medicare, Medicaid, and private insurance, covered participants in this program.

3. **Adults with chronic physical conditions—lower risk.** The six awardees in this group targeted individuals with chronic conditions, but who were not necessarily at high risk for high costs of care. The awardees in this group offered a diverse set of interventions, including the Community Care of North Carolina's pharmacy-focused program, technology-

supported mentoring, patient engagement for hepatitis C and HIV, and population health interventions. These awardees largely served Medicaid and Medicare beneficiaries.

4. **People (adults, young adults, or children) with behavioral health or cognitive disorders.** Five awardees served primarily individuals with behavioral health conditions or dementia. Amerigroup and Montefiore Medical Center focused on individuals with behavioral health challenges; the Clifford Beers Guidance Clinic targeted individuals with both behavioral and medical health issues. The remaining two awardees, Johns Hopkins University and the University of California at San Francisco targeted adults with dementia who were at risk of institutional care. All of these awardees included a care management and/or care coordination component in their programs, and most incorporated family engagement components. Montefiore Medical Center served individuals with public and private coverage. The other awardees served children covered by Medicaid or adults covered by Medicare or who are dually eligible.

5. **Adults who sought or who recently received acute or sub-acute care.** This group included six awardees that defined their target populations primarily by acute care episodes or transitions into and out of acute care. Although some participants had chronic conditions, this was not a criterion for enrollment. The awardees' programs generally served adults. Four of the six focused on Medicare and Medicaid beneficiaries, and the remaining two focused on individuals covered by a range of payers. A number of these programs provided transitional care coordination and therefore involved a range of settings, including medical respite care, care from 911 responders, and home care. However, the group also included awardees that offered pre-surgical education (University of Michigan) and tele-health for neuro-emergent conditions (University of New Mexico).

6. **People (adults, young adults, or children) who need primary or preventive care** but who do not meet the criteria for any of the other groups. The remaining eight awardees targeted populations in other ways, typically focusing on preventive care and/or access to specialized services. The programs in this group addressed oral health care, contraception, the risk of falling, and LBP. They also served age groups that differed greatly from one another, ranging from young children to adolescents to women at risk of pregnancy and to geriatric or other individuals with impaired mobility. The target populations for these awardees also brought a range of payers. There is no clear pattern of key innovations for these awardees, except that this group includes three of the six awardees who focused on extended provider roles.

C. Expected effects on core outcomes

The awardees also differed in the extent to which their programs focused on improving one or more of the four core outcomes identified by CMMI, as well as in the extent to which their programs will have an opportunity to affect these outcomes. The last four columns of Table II.1 show whether we expect, based on the awardees' theories of change and the effectiveness of program implementation, each program to have positive effects on the core outcomes.

Expected effects on core outcomes varied by awardee:

1. All-cause hospitalizations (25 awardees)
2. ED visits that do not lead to a hospitalization (26 awardees)
3. 30-day unplanned readmissions (20 awardees)
4. Total Medicare and Medicaid expenditures (32 awardees)

The vast majority of the awardees (32 of 38) designed and implemented their programs to reduce the total costs of care. The exceptions are the cases for which it is not clear whether costs will fall, but for which the innovation might improve quality of life outcomes (such as Amerigroup's program). Most of the awardees designed and implemented their programs to reduce all-cause hospitalizations, ED visits, and readmissions.

The variation in outcomes more commonly reflects the target population rather than the type of innovation. For example, programs that targeted high-risk individuals have had more of an opportunity to reduce the high utilization levels of this group, and the awardees designed their programs specifically to have such effects. Awardees that focus on preventive care, however, have had less of an opportunity to affect the core outcomes because of the short program period, and they tend to focus on improvements in care that are likely to affect longer-term outcomes.

In a few cases, the expected program effects differ from the four core outcomes used by CMMI. Altarum's focus on oral health is an example; Johns Hopkins University's program to delay or prevent nursing home admission for people with Alzheimer's and other dementia-related conditions is another. These awardees, as well as those reaching out to non-clinicians such as pharmacists and paramedics, are generally not expected to affect the core outcomes other than overall cost of care.

Table II.1. Association between target population, key innovation, and core outcome of HCIA R2 programs

Awardee by target population group	Key innovation						Core outcome			
	Care coordination	Patient education/ engagement	Care transitions	IT/ decision support	Tele-health	Extended provider roles	All-cause hospital admissions	ED visits	Hospital readmissions	Total cost of care
Youth with complex medical conditions										
BMC	X						X	X	X	?
NACHRI				X			X	X	X	X
SCH				X			X	X	X	X
UIC	X			X			X	X		X
WI DHS	X						X	X	X	X
High-risk chronic conditions										
NM			X		X		X	X	X	X
DMC	X						X	X	X	X
FSCL	X		X				X	X	X	X
Northwell	X	X	X				X	X	X	X
NYC H+H	X	X					X	X	X	X
U KS			X		X		X	X	X	X
UCSD	X						X	X		X
UHCMC		X		X						X
Lower-risk chronic conditions										
ACCF		X		X			?		X	X
CCNC	X			X			X	X	X	X
CHIIC							X	X	X	X
FPHNY	X				X		X			X
Ventura	X					X	X	X	X	X
VillageCare		X		X			X	?		X
Behavioral health and cognitive disorders										
Amerigroup	X							X		?
Clifford Beers	X	X					X	X	X	X
Hopkins	X		X			X	?	?	?	X
Montefiore	X	X		X			X	X	X	X
UCSF	X						X	X	X	X
Acute and subacute care										
CCC			X	X			X	X	X	X
Icahn	X		X				X	X	X	?
Mesa			X			X		X		X
NHCHC			X				X	X	X	X
UMich		X						?	?	X
UNM	X	X					X			X
Primary and preventive care										
AAMC			X	X				X		
Altarum						X				X
Avera				X	X		X	X		X
CHS	X							X		X
Columbia						X				X
U NC					X			?		?
Wash U		X					?	?		X
Yale	X					X	?	?	?	?

Source: Target population and key innovation are drawn from program summaries developed by the implementation research teams. Core outcomes have been populated by the implementation and impact research teams.

Notes: An '?' indicates that expected impact on core outcome is unclear.

Table II.2. Characteristics of HCIA R2 programs

Awardee	Program description	Key innovation	Target population
Altarum Institute (Altarum)	A multifaceted program to improve preventive dental care in pediatric primary care	Training and technical assistance to support primary care providers in conducting oral health screenings, applying fluoride varnishes, and referring patients to dental providers	Individuals age 17 and younger who were enrolled in Medicaid or CHIP, sought care from participating primary care providers, and resided in Michigan at the time of the study
American College of Cardiology Foundation (ACCF)	Implementation of SMARTCare, a combination of clinical decision support, shared decision-making, patient engagement, and provider feedback tools designed to improve care for patients with stable ischemic heart disease	Bundling decision-support tools based on evidence-based medicine; bundled approach to paying for services	Individuals who were enrolled in Medicare or had private insurance, had stable ischemic heart disease, and resided in Florida or Wisconsin
Amerigroup	Provided coaching services for youth transitioning out of foster care	Wraparound services: medical, educational, employment, and social; care coordination	Youth ages 17 to 20 who were enrolled in Medicaid and were in foster care, had a documented history of behavioral health needs, and resided in participating counties in Georgia
Association of American Medical Colleges (AAMC)	Tested the scalability of eConsult, an enhanced referral model, to five academic medical centers	Formalized communication between primary care providers and specialists, leverage the EMR system to facilitate communication, expand access to specialists' feedback	All individuals age 17 and older who visited participating primary care practices in California, the District of Columbia, Illinois, Iowa, New Hampshire, Virginia, or Wisconsin
Avera Health (Avera)	Virtual wrapping of a set of comprehensive, resident-centered, geriatric care services	Use of video and audio technology to provide instant, round-the-clock access to health care providers; development and use of a risk stratification algorithm for the geriatric population in skilled nursing facilities	All individuals who resided in participating SNFs in Iowa, Minnesota, Nebraska, and South Dakota
Boston Medical Center (BMC)	Paired complex care nurse coordinators and complex care pediatricians with community primary care providers to enhance care for children with medical complexity in medical home-like settings	Using interdisciplinary care teams made up of social workers, family navigators, complex care pediatricians, nurse care coordinators, and specialists; intensive and personalized care plans for each patient	All children with medically complex conditions who resided in Massachusetts. Medically complex conditions were defined as having high utilization in the year before referral or clinical assessment of being at risk of high utilization.
CareChoice Cooperative (CareChoice)	Used a web-based application that incorporates components of the hospital-based Re-Engineered Discharge program adapted to nursing home transitional care units	Used additional staffing (hiring of transition coordinators), decision-making tools (Engage software), and process improvement activities to improve the quality of care and safety provided to nursing home residents discharged back to the community	Individuals who were enrolled in Medicare or Medicaid or who were dually enrolled, who were admitted to a participating nursing facility as a transitional care unit patient, and who resided in Minnesota

Table II.2 (*continued*)

Awardee	Program description	Key innovation	Target population
Catholic Health Initiatives Iowa Corp. (Catholic Health Initiatives)	Transitioned a network of rural critical access hospitals and their affiliated practices to value-based care through improved chronic disease management, population health activities, and lean process improvement initiatives	Expanded existing urban-based population health activities to Medicare and Medicaid beneficiaries in resource-limited rural communities. In addition, the awardee was an early proponent of bringing rural practices into a Medicare Shared Savings Program.	Adults enrolled in Medicare or Medicaid with one or more chronic diseases who sought care at one of the Mercy Health Network–affiliated clinics in rural areas in northern, central, or western Iowa and eastern Nebraska
Children's Home Society of Florida (CHS)	Implementation of a medical home to reduce ED and inpatient utilization, increase awareness of sexually transmitted disease, and address food insecurity and stress	Developed an intervention within an existing Community Schools model to expand access to health-related services for the students and others affiliated with the school and to the community at large. The expansion consisted of (1) patient navigators to connect students and community members with needed services and (2) direct service providers (including medical, dental, and behavioral health) at the Community School site.	All individuals who would benefit from patient navigation or direct medical, dental, or behavioral health services and who resided in Pine Hills, Florida
City of Mesa Fire and Medical Department (Mesa)	Two components: The first offered on-site evaluation and treatment through mobile community medicine units for low-acuity patients who used the 911 system and the ED; the second followed high-acuity patients at high risk of readmission after hospital discharge	Used mobile community medicine units staffed with either an advanced practice provider and a paramedic or a licensed behavioral health clinician and a paramedic to provide direct care to low-risk 911 callers and conducted home visits and care coordination for high-risk patients within 72 hours of hospital discharge	For the first component, all individuals who called 911 about low-risk concerns. For the second component, patients with congestive heart failure, COPD, heart attack, sepsis, or pneumonia who were recently discharged from Mountain Vista Medical Center or Dignity Health who were identified as high-risk for readmission and resided in the Mesa Fire and Medical service area
Clifford Beers Guidance Clinic (Clifford Beers)	Provided services to improve care coordination, increase patient/family engagement, and reduce fragmentation of services and the total cost of care	Wraparound care coordination services, family engagement	Children enrolled in Medicaid with complex physical and behavioral health needs who resided in the New Haven area of Connecticut
Community Care of North Carolina (CCNC)	Integrated medication management strategies provided by community pharmacies into existing patient-centered care teams while incentivizing pharmacists to address gaps in care	Engaged and incentivized pharmacists through a value-based approach that changes their role from simply filling and dispensing medication to providing enhanced services	Individuals enrolled in Medicaid FFS, Medicare, CHIP, or who are dually enrolled; all of these beneficiaries must also have had one or more chronic medical conditions treated through medications filled at participating pharmacies in North Carolina
Trustees of Columbia University in the City of New York (Columbia)	Community health workers used technology-assisted behavioral risk-reduction strategies to reduce early childhood caries in low-income children	Used a chronic disease management model within pediatric dentistry to improve the experience of care and reduce costs	Children ages 2 to 6 who were enrolled in Medicaid (plus up to two of their siblings and their caregivers), who had early childhood caries and no comorbidities, and who resided in New York City

Table II.2 (*continued*)

Awardee	Program description	Key innovation	Target population
Detroit Medical Center (DMC)	Established patient-centered medical home clinics adjacent to EDs to increase availability of primary care for individuals who arrived at the ED and required non-urgent care	Patient-centered medical home primary clinics embedded in the ED	All individuals who were frequent users of the ED, had no primary care physician on record, had at least one of seven chronic conditions (defined as diabetes, asthma, hypertension, congestive heart failure, depression, chronic obstructive pulmonary disease, or HIV/AIDS), and resided in Detroit, Michigan
Four Seasons Compassion for Life (FSCL)	Community-based palliative care model that featured interdisciplinary collaboration and the integration of palliative care into the health care system; continuous access to care during transitions from one care setting to the next; and longitudinal, individualized support for patients and families	People with life-limiting illnesses received community-based, patient-centered palliative care through a multidisciplinary care team that offered psychosocial and spiritual support as well as clinical care in both inpatient and outpatient settings. The innovation also educated providers and the public about the benefits of palliative care.	Adults older than 65 who were enrolled in Medicare FFS, had a life-limiting illness (usually with a prognosis of one year or less), and resided in western North Carolina or in the Greenville area of South Carolina
Fund for Public Health in New York, Inc.(FPHNY)	Identified and treated people with hepatitis C virus	Improved treatment for the hepatitis C virus in New York City by using patient-centered medical and behavioral health care integrated with care coordination and tele-mentoring	Adults enrolled in Medicare or Medicaid who were infected with the hepatitis C virus and who resided in the Bronx or in the East or Central Harlem areas of New York City
Icahn School of Medicine at Mount Sinai (Icahn)	Tested Mobile Acute Care Team (MACT) services to address acute care needs in the home setting	Expanded the Hospital at Home payment model to provide additional services, such as sub-acute rehabilitation, to different populations	Individuals at least 18 years of age who were enrolled in Medicare FFS or Medicaid managed care, or were dually eligible who presented in Mount Sinai inpatient and outpatient settings and were discharged for acute and sub-acute rehabilitation at home and resided in Manhattan
Johns Hopkins University	Alzheimer's disease/dementia-targeted care coordination model	Home-based dementia care coordination and caregiver support, including recurring home visits conducted by memory care coordinators	Older adults (and their caregivers) who were dually enrolled in Medicaid and Medicare or in Medicare only, had Alzheimer's disease or a related form of neurodegenerative dementia, and resided in the Baltimore, Maryland area
Montefiore Medical Center (Montefiore)	Collaborative care management to address serious behavioral health needs	Implemented an existing integrated care model (the Collaborative Care Model) with a high-need population; used technology platforms to improve patient engagement and service delivery	Individuals with public or private health insurance who received services from participating primary care sites; screened positive for depression, anxiety, or (for children and adolescents) attention deficit hyperactivity disorder; and resided in the Bronx

Table II.2 (*continued*)

Awardee	Program description	Key innovation	Target population
National Association of Children's Hospitals and Related Institutions (NACHRI)	National learning collaborative for programs targeting children with medical complexity	Learning collaborative that included hospital-based practices and primary care practices across multiple states to implement a package of interventions tailored to local needs to improve care for CMC	Children who were enrolled in Medicaid with a long-term chronic condition, a complex chronic condition, or a malignancy (as defined by 3M Clinical Risk Groups software categories 5b, 6, 7, 8, or 9 using billing or claims data) and who resided in states that have participating hospitals
National Health Care for the Homeless Council (NHCHC)	Medical respite care for homeless Medicaid and Medicare beneficiaries following discharge from a hospital or other community-based setting	Implemented a standardized model of respite care in five sites. The model was defined by the delivery of a core set of services representing recognized best practices in the field of respite care.	Individuals age 18 and older who were enrolled in Medicare or Medicaid or who were dually enrolled and who experienced homelessness, had an acute illness or injury, were at high risk of hospitalization, and resided in the vicinity of a participating respite care program in Arizona, Connecticut, Minnesota, Oregon, or Washington State
Nebraska Medical Center (NM)	Remote patient monitoring for 90 days post-discharge using tele-health consultations	Tele-health for clinic visits with patients 90 days after the end of patient monitoring	Adults with Type 2 diabetes were recently discharged from the hospital, were at high risk for readmission, and resided in medically underserved areas (defined by zip code) in the Omaha area of Nebraska
New York City Health + Hospitals Corporation (NYC H+H)	An ED care management model to improve the linkage to primary care and preventive health services, and to reduce utilization of the ED for ambulatory care sensitive conditions	An ED-based team of nurse care managers, pharmacists, community liaison workers, and home care intake nurses helped patients to better manage their health by providing education, support, and linkages to ambulatory care and home health care as needed	All adults who visited an ED for an ambulatory care sensitive condition or who met other utilization-based criteria (such as another recent ED visit or hospitalization) and resided in New York City
Northwell Health (Northwell)	Provided education to and coordinated care for patients with advanced CKD in order to improve decision making, increase quality of life, and reduce total cost of care through reduced hospitalization and ED use	Focused on transition between advanced CKD and end-stage renal disease. to address progression by planning for fistula placement and dialysis options	Adults who had advanced CKD and resided in one of four counties in the New York City area
University of California at San Diego (UCSD)	Implemented the Heart Attack and Stroke Free-Zone project by increasing awareness of risk factors, promoting evidence-based practices, and testing technology solutions	Introduced and provided education on evidence-based medication bundles to patients and providers in conjunction with supportive, ongoing health coaching to patients	Individuals who were enrolled in Medicaid or Medicare or who were dually eligible; at high risk for a major adverse cardiovascular event (such as a heart attack, stroke, or sudden cardiac death); and resided in San Diego, California

Table II.2 (*continued*)

Awardee	Program description	Key innovation	Target population
University of California at San Francisco (UCSF)	Provided care management and caregiver support for participants with dementia to improve patient/family satisfaction with care, reduce caregiver burden, prevent emergency-related health care costs, and keep patients in the community longer	Care management and caregiver support provided by phone. The awardee designed a randomized controlled trial.	Individuals age 45 and older (and their caregivers) who were enrolled in Medicare or Medicaid with a diagnosis of dementia and resided in California, Nebraska, or Iowa
University Hospitals Cleveland Medical Center (UHCMC)	Provided patient-centered care management and coordination to cancer patients to improve quality of care and patient satisfaction, and to reduce the total cost of care and demonstrate the feasibility and sustainability of a new shared savings payment model	Improved the experience and appropriateness of care by promoting evidence-based practices such as early palliative care, nurse care coordination, and spiritual care. Program staff administered a biopsychosocial assessment to all patients via an iPad to collect patient-reported data on outcomes and used the data to engage patients in their care and to formulate care plans.	Adults who were enrolled in Medicare or Medicaid, received care at Seidman Cancer Center for complex cancers, and resided in Ohio
University of New Mexico (UNM)	Expanded the existing tele-health infrastructure to form a statewide hospital tele-system	Tele-health services for remote monitoring of patients with diabetes	Adults enrolled in Medicare or Medicaid who presented with a neuro-emergent condition at a participating ED in New Mexico
University of Illinois	Implemented the CHECK program, a medical neighborhood via a network of over 40 practices	Deployed community health workers to build relationships with children and their families, identified health and social service care coordination needs, and connected participants to information and services that addressed these needs. The CHECK program's health IT provided resources and communication support to community health workers and allowed for remote access to mental health services.	Children and young adults age 25 and younger who were enrolled in a Medicaid managed care plan or in Medicaid FFS; had chronic medical conditions (defined as being diagnosed with diabetes, sickle cell disease, asthma, or prematurity); and resided in Cook County, Illinois
University of Kansas Hospital Authority (U KS)	Rural, clinically integrated network of providers focused on improving heart health and survival after stroke by (1) implementing acute care protocols for patients presenting with signs and symptoms of heart attack or stroke, (2) developing tele-health solutions, (3) implementing regional transitional care management, and (4) providing chronic care management	U KS formed a patient safety organization and developed a rural health payment model. It also used remote critical care staff to support Avera eCare emergency consultation.	All residents who were hospitalized for heart attack, stroke, or sepsis, were at risk for heart attack or stroke, or had hypertension or hyperlipidemia, and resided in Kansas
University of Michigan (UMich)	Optimized pre-operative care for patients undergoing major inpatient surgery by enrolling them in a patient education and physical activity monitoring program	Engaged pre-surgical patients through a patient education kit with a pedometer, a spirometer, and an automated program to remind patients to log the number of steps they walked at least one week before surgery	All individuals who were scheduled for a major abdominal or a select general surgery at participating surgical practices throughout Michigan and were at high risk for poor surgical outcomes

Table II.2 (*continued*)

Awardee	Program description	Key innovation	Target population
University of North Carolina (U NC)	Care delivery model for new onset LBP; model included patient education and shared decision-making tools, and nurses who were patient navigators	A checklist for LBP, which was integrated into the EMR or provided on paper. The checklist (1) prompted participating providers to follow an evidence-based treatment protocol for all patients presenting with new, acute LBP and (2) offered decision support.	Individuals enrolled in Medicare and Medicaid who visited a participating outpatient primary or specialty care provider for acute nonspecific LBP, had not seen a provider for back pain in the previous six months, and resided in a seven-county region in North Carolina
Seattle Children's Hospital (SCH)	Provided care management and coordination for children with complex health conditions to improve health outcomes, reduce medical costs, and develop a scalable outpatient care management model	The program was a new approach in the region toward care coordination for children with complex needs. It was innovative in its tight relationships with Medicaid managed care organizations with respect to the development of services and the payment model.	Children enrolled in both Medicaid and the Supplemental Security Income program, were identified as being at high risk for negative health outcomes, and resided in King or Snohomish counties in Washington State
Ventura County Health Care Agency (Ventura)	Community-based care coordination program for patients with COPD	Care coordination for early-stage patients with COPD (or risks for COPD) that had a high quality, validated staging and treatment regime (the GOLD standard). Early identification and appropriate treatment together can help improve lung functioning, along with the overall quality of life, and reduce adverse advents such as unplanned hospitalizations.	Individuals enrolled in Medicare or Medicaid who had COPD and resided in Ventura County, California
VillageCare	Technology-based self-management tool to increase patient activation and treatment adherence for people living with HIV	Used a technology-based self-management tool to improve adherence to HIV/AIDS treatment	Individuals age 18 and older who were enrolled in Medicare or Medicaid, were diagnosed with and prescribed medication for HIV, and resided in the New York City area
Washington University School of Medicine in St. Louis (Wash U)	Family planning services for women who were at the highest risk for unintended pregnancy	Improved access to effective methods of contraception through the use of insurance navigation, comprehensive contraceptive counseling and support, and same-day access to contraceptive services, including long-acting reversible methods	All women ages 14 and older who were at high risk for unintended pregnancy and childbirth and resided in the St. Louis area of Missouri
Wisconsin Department of Health Services (WI DHS)	New pediatric provider and reimbursement models for the care of children with medical complexity and high-resource utilization	WI DHS and its hospital partners implemented the Special Needs Program (SNP) to test a pediatric provider and reimbursement model for the care of CMC whose families frequently faced disjointed care and significant stress. Through the SNP, WI DHS sought to address the needs of CMC and their families by providing integrated health care, including direct and consultative patient care, care management, and care coordination.	CMC with fragility who were enrolled in Medicaid or CHIP and resided in Wisconsin. Complex medical conditions are defined by chronic conditions involving three or more organ systems that require ongoing care from three or more specialists. Fragility is defined as high tertiary center utilization.

Table II.2 (*continued*)

Awardee	Program description	Key innovation	Target population
Yale University	Expanded paramedic workforce with advanced care training to improve care coordination and health outcomes for elders staying in their homes	Provided elders and individuals who have impaired mobility with in-home interventions and increased linkages with primary care in order to reduce fall risk	All individuals who were living at home and have fallen or were at risk of falling and resided in the emergency medical services geographic catchment area of Greater New Haven, Connecticut

Source: "Program description," "key innovation," and "target population" are drawn from program summaries developed by the implementation research teams.

CHIP = Children's Health insurance Program; CKD = chronic kidney disease; CMC = children with complex medical conditions; COPD = chronic obstructive pulmonary disease; ED = emergency department; EMR = electronic medical record; FFS = fee-for-service; HIV = human immunodeficiency virus; LBP = lower back pain; SNF = skilled nursing facility.

III. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Introduction and summary of findings

This chapter presents a synthesis of the findings from the implementation evaluations of the 38 awardees. The synthesis has three main objectives. The first is to assess the effectiveness of program implementation. The second is to identify the lessons learned from the implementation of the cooperative agreements, including (1) the factors that maximize the chances of implementation effectiveness, (2) the barriers that prevented some awardees from reaching full implementation effectiveness, and (3) the strategies that awardees of effectively implemented programs used to overcome common problems. The third objective is to highlight the implications of awardees' implementation performance for the interpretation of the impact results. It should be emphasized that the ultimate objective of this assessment was not to grade awardees on their implementation performance. Rather, the objective was to create two awardee groups based on the degree of implementation effectiveness and then to use these groups to identify the factors associated with that success.

The key findings from the implementation evaluation include the following:

- More than half of the awardees (20) met at least 90 percent of their final enrollment projections (set at the end of Year 2). Another one-third (13) met 65 to 90 percent of their final projections. Five awardees failed to meet 65 percent of their final enrollment targets.
  - Awardees could, and most did, revise their enrollment goals during the cooperative agreement. The 24 awardees that lowered their initial goals were more likely to meet their final enrollment projections than were the 14 that either increased their projections or left them unchanged during the cooperative agreement.
  - When measured against the enrollment goals set by awardees at the end of Year 1, enrollment performance was weaker. Only 14 awardees met at least 90 percent of their end-of-Year-1 goals, 16 met 65 to 90 percent, and 8 failed to meet 65 percent of their goals.
- We concluded that 24 of the awardees implemented their service delivery models effectively. That is, any problems that the awardee might have encountered with enrolling participants, delivering services, employing and training staff, and engaging providers and patients were not deemed serious enough to threaten the awardee's ability to improve its targeted outcomes. The other 14 awardees were partly effective in implementing their service delivery models, implying that shortcomings in one or more of these areas might have impeded their ability to improve outcomes.
- Programs with these structural features were more likely to have been implemented effectively: (1) targeted chronic rather than acute diagnoses, (2) focused on changing providers' rather than patients' behaviors, (3) served single or local areas rather than large or multiple areas, and (4) had piloted-tested their model rather than implemented an untried program.

- The evaluation team identified six activities or situations associated with effective service delivery:
  - **Creating a supportive internal environment** through a team-oriented culture, strong communication and supportive structures, and the presence of program champions
  - **Obtaining the active participation of providers and community organizations** by understanding their constraints and needs, and by offering services perceived as valuable
  - **Keeping participants engaged in the program** by building trust and rapport with them through personal hand-offs between clinicians and staff and open communication via texting, tailoring services to participants' needs, and furnishing services perceived as valuable
  - **Attending to ongoing staffing issues** by hiring, training, and retaining enough staff with the appropriate skills for the program
  - **Deploying health information technology** (health IT) to facilitate the use of clinical data in patients' care, support the flow of patients' data among providers and between providers and patients, and offer patients ways to engage with and manage their own care
  - **Having external circumstances favorable to the program,** such as regulations and policies aligned with program goals and an adequate supply of health care resources

The remainder of this chapter is divided into five sections. Following the introduction and summary of key findings, **Section B** describes the data and methods used to assess implementation effectiveness. **Section C** presents the results of our implementation effectiveness assessment. **Section D** provides an assessment of the perceived effects of the HCIA R2 programs on the delivery of care and health outcomes, based on responses to the staff and clinician surveys. In **Section E**, we identify the factors influencing implementation effectiveness around enrollment and service delivery models and the strategies used to overcome common challenges. **Section F** highlights the implications of program implementation for the design and interpretation of the impact evaluations.

B. Data and methods for assessing implementation effectiveness

1. Data sources

The findings presented in this chapter are based on the evaluation team's analyses of three data sources.

- **The awardees' self-reported information** submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. The information included the awardees' operating plans, self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials.
- **Qualitative information collected annually through in-person and virtual site visits** toward the end of each of the three years of the cooperative agreements. We used semi-

structured protocols tailored to the features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites.

- **Responses to two surveys—one of non-clinician staff and one of clinicians**—on their perceptions of the effect of the programs on the delivery of care and health outcomes.[2]
  - The **non-clinician staff survey** (staff survey hereafter) was fielded to health care staff from 35 of the 38 awardees from July to October 2016. Two awardees did not receive the survey because very few staff were involved in the program (the University of North Carolina and the American Association of Medical Colleges), and one awardee could not provide a staff list for its sites because of internal institutional review board (IRB) concerns (the National Association of Children's Hospitals). We defined staff as non-clinician employees who were involved in program implementation, aspects of participants' care, or services. Examples of the types of staff are social workers, care coordinators, registered nurses, medical assistants, patient navigators, and health IT staff. The overall adjusted response rate for the staff survey was 79.9 percent, with a total of 914 completes. Three awardees had a response rate below 67 percent, and six had a response rate of 100 percent (see Appendix A). The number of completed interviews for individual awardees ranged from 5 to 148.
  - The **clinician survey** was fielded to clinicians at 18 of the 38 programs from March to June 2017. We did not collect data from 20 awardees, mainly because of limited or lack of clinician involvement in the program or because there were too few clinicians to allow reporting. Clinicians included physicians, dentists, nurse practitioners, and physician assistants who referred patients to the programs, directly provided care or services to participants, or provided consultation and support to staff or other clinicians who directly provide services to program participants. The overall adjusted response rate for the clinician survey was 63.3 percent, with a total of 1,002 completed responses. Six awardees had a response rate below 67 percent, and four had a response rate of 100 percent (see Appendix A). The number of completed clinician surveys for individual awardees ranged from 4 to 167, reflecting wide variation in the size of the awardees' staff.

2. Methods for assessing implementation effectiveness

To assess implementation effectiveness, we considered each awardee's program enrollment and its success in implementing the service delivery model on four domains listed in Table III.1. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on (1) **the delivery of intervention services**, (2) **staffing and training**, (3) **recruitment and engagement of providers**, and (4) **engagement of participants**. Table III.1 describes the criteria used to assess

[2] A detailed description of the survey methodology and survey instruments is available from the authors on request. The numbers of completed surveys and response rates for each awardee that received a staff or clinician survey are provided in Appendix A.

implementation effectiveness for program enrollment and each of the four service delivery domains.

Table III.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in these features affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful and sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of its participants for the full period of enrollment (as applicable)?

[a] Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; 65 to 90 percent = partly achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness. Awardees were allowed to change their enrollment projections (up or down) at the end of the first and second years of their cooperative agreements. We based our assessment of enrollment effectiveness on awardees' final enrollment count relative to their projections as reported in their 9th quarterly reports (end of Year 2).

To conduct this assessment, we first looked at enrollment and scored the awardees' performance mainly on the percentage of their three-year cumulative enrollment projections met. We considered awardees that met at least 90 percent of their final three-year projections to be effective, those that met 65 to 90 percent of their projections as partly effective, and those that enrolled less than 65 percent of their final projections as ineffective. We chose these thresholds based on the distribution of the percentage of projections met across the programs.

We then looked at the four service delivery domains and applied the criteria listed in Table III.1 to each program. We considered programs that met most or all of the criteria for a given domain to have been implemented effectively, and we considered those that did not meet the criteria to have been implemented partly effectively. To be considered implemented effectively in the service delivery component, an awardee had to meet the criteria in the delivery of services domain (domain #1 in Table III.1) plus (in most cases) the criteria for at least two other domains. However, the number of other domains needed for a program to be implemented effectively could vary depending on the design of the program. For example, the engagement of patients may not have been relevant for practice-level interventions in which awardees use passive enrollment; we note in the table when a domain was not applicable to the intervention. One

awardee was determined to be effective in staffing and in provider and patient engagement but failed to meet the threshold for full effectiveness in the delivery of services. We based our assessment mainly on qualitative information collected through document review and interviews, and corroborated to the extent possible with the survey responses. We created a composite service delivery score by looking at the preponderance of evidence across the four domains.

Finally, we created a score of overall implementation effectiveness based on a combination of the enrollment score and the composite service delivery score. Each awardee received either an overall effective score or an overall partly effective score. For 22 of the 38 awardees, our scores for enrollment and service delivery were the same. That is, we scored these awardees as "effective" or "partly effective" for both the enrollment and service delivery domains. Of the 16 awardees for which our scores differed, we based our overall assessment on the service delivery score because we considered success in delivering intervention services to be more important to implementation effectiveness than meeting a sometimes unknowable and changing enrollment target. As previously noted, we also lacked information to consistently assess the timing of enrollment, the duration of exposure, and awardees' success in enrolling patients most likely to benefit from the intervention—issues that will be more fully explored in the impact analysis.

C. Results of the implementation effectiveness assessment

In this section, we present the findings of our implementation effectiveness analysis. We start by presenting the results of the enrollment and service delivery assessments separately, and then discuss the overall implementation effectiveness scores.

1. Effectiveness in enrolling program participants

Differences in how awardees recruited participants, defined enrollment, and adjusted their enrollment projections during the cooperative agreement make assessing program enrollment effectiveness inherently challenging. Some awardees required patient consent and actively enrolled patients into the program (called active enrollment). Other awardees served everyone who sought care at a participating site who met the awardee's eligibility criteria, many of whom but did not know they were enrolled (called passive enrollment). A second important distinction is whether the people who benefited from the program, whom we refer to as participants, received person-specific services directly funded by the awardee's HCIA R2 cooperative agreement (such as care coordination or patient education and counseling) or whether they simply benefited from broad practice-level changes (such as in workflow improvements and investments in new or upgraded technologies) that the awardees implemented under their agreements to benefit all of their patients indirectly. Some awardees served their participants both directly and indirectly—providing services directly funded by the cooperative agreements and making HCIA R2–funded structural changes that affected all patients. Participants can receive services directly provided by the program (for example, case management) regardless of whether the awardee actively or passively enrolled people. Receipt of indirect services (called indirect participation) occurred mainly for awardees that passively enrolled people.

In addition to defining enrollment differently, awardees were allowed to revise their enrollment projections in the fifth and ninth program quarters of their cooperative agreements. The original projections served as an enrollment *target* or *goal*, based on awardees'

understanding of the need for program services at the beginning of their programs. The revised counts often reflected the number of participants whom the awardees expected to serve. For consistency, we use the term *projections* to mean both the awardee's original goal as well as the final number of participants the awardee expected to serve by end of the program.

More than half (20) of the 38 awardees met at least 90 percent of their final (often revised) enrollment projections by the end of their initial three-year cooperative agreements. Fourteen of these 20 awardees met or exceeded their projections (see Table III.2). One-third of the awardees (13) met 65 percent to 90 percent of their final projections. Only 5 awardees failed to reach at least 65 percent of their final projections, one of which met less than half of its projection.

Two awardees increased their projections by over 1,000 percent in Year 3. Community Care of North Carolina based its enrollment numbers on patients attributed to participating pharmacies, as opposed to participants receiving intervention services. The network of participating pharmacies grew faster than anticipated. As a result, enrollment numbers increased faster than anticipated as well. Catholic Health Initiatives raised its projection of direct participants several times during the three-year cooperative agreement because it felt that its initial target was too low. In the end, the awardee far exceeded its original target but met only two-thirds of its revised three-year projection. We therefore scored the awardee as being only partly effective in enrollment.

The number of awardees that met their enrollment target by the end of Year 3 depends on projection year used to make this determination.

Percent of enrollment target met based on projections from:	> 90 percent	65-90 percent	<65 percent
End of Year 1	14	16	8
End of Year 2	20	13	5

Table III.2. Changes in participant projections and percentage of final projections met, sorted by percent of target met

Awardee	3-Year participant projections at beginning of Year 1	3-Year participant projections at beginning of Year 3	Percent change in 3-year projections from Year 1 to Year 3	Number of participants served from program inception through end of Year 3	Percentage of 3-year projections met through end of Year 3
Avera	5,521	4,134	-21.4	6,778	164.0
Altarum	1,000,000	742,715	-25.7	949,164	127.8
Northwell	593	500	-15.7	629	125.8
WI DHS[a]	1,360	721	-47.0	905	125.5
Amerigroup[b]	720	720	0.0	860	119.4
Yale	4,800	3,700	-22.9	4,288	115.9
Hopkins	600	300	-50.0	342	114.0
UCSF	1,400	458	-67.3	512	111.8

Table III.2 *(continued)*

Awardee	3-Year participant projections at beginning of Year 1	3-Year participant projections at beginning of Year 3	Percent change in 3-year projections from Year 1 to Year 3	Number of participants served from program inception through end of Year 3	Percentage of 3-year projections met through end of Year 3
ACCF	84,583	26,832	-68.3	29,053	108.3
FSCL	8,000	5,371	-32.9	5,803	108.0
AAMC	155,220	120,000	-22.7	128,721	107.3
UIC	6,000	8,200	36.7	8,455	103.1
U NC	10,662	1,459	-86.3	1,472	100.9
NACHRI	9,401	8,064	-14.2	8,111	100.6
U KS[c]	7,739	7,363	-4.9	7,334	99.6
UCSD	4,008	3,800	5.2	3,731	98.2
FPHNY	3,200	2,914	-8.9	2,775	95.2
Montefiore	4,575	6,380	39.5	6,060	95.0
CCNC[d]	27,954	356,853	1,176.6	328,806	92.1
CCC	12,057	8,874	-26.4	8,016	90.3
VillageCare	5,121	5,036	-1.7	4,366	86.7
Ventura	2,500	2,500	0.0	2,162	86.5
Mesa	27,179	14,543	-46.5	12,417	85.4
DMC	10,500	8,200	-21.9	6,996	85.3
Clifford Beers	2,250	2,284	1.5	1,944	85.1
SCH	1,600	960	-40.0	813	84.7
NYC H+H	107,706	100,174	-7.0	83,946	83.8
BMC	450	450	0.0	365	81.1
CHS	5,009	7,481	49.4	6,017	80.4
NM	3,300	2,500	-24.2	1,903	76.1
UHCMC	1,776	1,776	0.0	1,331	74.9
Wash U	10,000	4,047	-59.5	3,022	74.7
CHIIC[e]	1,295	20,000	1,444.4	13,044	65.2
Columbia	1,936	1,936	0.0	1,207	62.3
UNM	8,504	1,968	-76.9	1,235	61.8
UMich[f]	10,000	2,000	-80.0	1,235	61.7
Icahn	1,080	1,080	0.0	605	56.0
NHCHC	2,890	3,127	8.2	1,378	44.1

Source: Participant projections at the beginning of Year 1 were obtained from the first program quarter (September 2014–November 2014) data provided by the implementation and monitoring contractor except where noted. Participant projections at the beginning of Year 3 were obtained from the 9th program quarter (September 2016–November 2016) data provided by the implementation and monitoring contractor except where noted. The number of participants served was obtained from 12th program quarter (June 2017–August 2017) data provided by the implementation and monitoring contractor.

Notes: The participation counts include all participants regardless of payer. The data on the number of participants are self-reported by the awardee and have not been verified by Mathematica. For most awardees, the results are based on direct participants only (including the 10 awardees that had both direct and indirect participants). For the 7 awardees that had only indirect participants, we measured participants relative to the awardees' three-year projections by using indirect participants only. We show in *italics* the 17 awardees that served at least some participants indirectly. In **bold**, we show the 13 awardees that used passive enrollment.

Table III.2 *(continued)*

[a]In September 2016, WI DHS revised its original 3-year projections downward to 1,121 participants from the 1,470 participants set in June 2016. However, the awardee made a data entry error while updating the projection reported to the implementation and monitoring contractor in the 9th program quarter. WI DHS' projection includes both direct and indirect participants. For the purposes of this report, we present the direct and indirect participant projections reported to the implementation and monitoring contractor.

[b]In reports to the implementation and monitoring contractor, Amerigroup classified its participants as indirect. Because these participants receive HCIA R2–funded services directly, we reclassified Amerigroup's participants as direct participants in this report.

[c]In the 11th program quarter (May 2017), U KS revised its enrollment target to 7,363 indirect participants. The previous target submitted by U KS to the implementation and monitoring contractor at the beginning of Year 3 was based on 17,070 indirect enrollees, not on indirect participants served, which is the appropriate target for U KS.

[d]The total number of participants served and the percentage of three-year projections met are inflated for CCNC. They reflect the number of participants who sought any service at participating pharmacies, not the total number of participants who received HCIA R2–funded services directly from a participating pharmacy. We do not have an estimate of the number of participants served under the program.

[e]CHIIC increased its participant projections several times after the program was launched because it realized that its initial goal was too low. In addition, the awardee stated that the final participant count under-reports total participation. Participating clinics in one of the three regions transitioned to a new data collection system in Year 2, and were unable to integrate the new system with the existing program tracking software.

[f]UMich lowered its 3-year direct participant projections from the original goal of 10,000 participants twice, first to 4,194 in March 2016 and again to 2,000 in August 2016. UMich defined direct participants as unique participants who had either Medicare or Medicaid as their primary insurer and indirect participants as those who did not have Medicare or Medicaid as their primary insurer.

Nine of the 13 awardees (69 percent) that used passive enrollment reached or nearly reached their final projections, compared with only 10 of the 25 awardees (40 percent) that used active enrollment (Table III.3). Participants included all individuals who saw a program provider and met the program eligibility requirements. Awardees that used passive enrollment also tended to have higher enrollment projections than those that used active enrollment. Awardees that served participants indirectly were also more likely to have met at least 90 percent of their final projections than those that served participants only directly. Six of the 7 awardees that relied on passive enrollment and that only served patients indirectly met at least 90 percent of their final projections, and the seventh awardee exceeded 65 percent.

Table III.3. Enrollment effectiveness, by type of enrollment and participation

Type of enrollment	Type of participation	Number of awardees	Number (percentage) of awardees who met:	
			At least 90 percent of final three-year projections	Less than 65 percent of final three-year projections
Passive	Direct only	3	1 (33%)	1 (33%)
	Indirect only	7	6 (86%)	0 (n.a.)
	Direct and indirect	3	2 (67%)	0 (n.a.)
Active	Direct only	18	8 (44%)	3 (17%)
	Indirect only	0	0 (n.a.)	0 (n.a.)
	Direct and indirect	7	2 (29%)	2 (29%)

Source: Participant projections were obtained from the 9th program quarter (September 2016–November 2016) data provided by the implementation and monitoring contractor. The number of participants actually served was obtained from the 12th program quarter (June 2017–August 2017) data provided by the implementation and monitoring contractor.

Note: A direct participant is an individual who received care or services that were not currently covered by Medicare or Medicaid, but were paid for directly by the HCIA R2 program funding, such as care

Table III.3 *(continued)*

coordination services, patient education and counseling, or medication management. Some awardees required direct participants to actively enroll in their program, while other awardees did not. An indirect participant was anyone who benefitted from HCIA R2 funding, due to enhancements made at the provider level (such as clinician training or investments in health information technology) rather than to receiving supplemental services paid for with HCIA R2 funding.

n.a. = not applicable.

However, these findings on enrollment provide an incomplete and potentially misleading picture of enrollment effectiveness for five reasons. First, before the awardees began to enroll beneficiaries, they were required to report the number of participants they expected to serve by the end of the cooperative agreement. Some awardees could not make an accurate prediction because they did not know how many people would be eligible for their programs. As mentioned, CMS gave awardees the opportunity to revise their projections in the fifth and ninth program quarters. Thirteen of the 20 awardees (65 percent) that *decreased* their projections in the ninth program quarter by 10 percent or more met at least 90 percent of their projections, compared with only 4 of the 13 awardees (31 percent) that stayed within plus or minus 10 percent of their projections (Table III.4). Nearly a quarter (23 percent) of the awardees that either did not change their projections or stayed within plus or minus 10 percent of their original projections struggled to meet their enrollment goals. When measured against the enrollment projections set by awardees at the end of Year 1, enrollment performance is weaker. (This interval allowed awardees enough time to use knowledge gained during Year 1 to adjust their initial goal to a more realistic level, but was set early enough to still be a goal rather than a projection of what was likely to happen in the final year.) Only 14 awardees met at least 90 percent of their projections from the end of Year 1, 16 awardees met 65 percent to 90 percent of their projections, and 8 awardees failed to meet 65 percent of their projections.

Enrollment findings provide an incomplete and potentially misleading picture of enrollment effectiveness because:

1. Awardees were allowed to change enrollment projections during cooperative agreement
2. Some awardees changed eligibility criteria in ways that likely weakened program impacts
3. Measure does not assess how successful awardees were in enrolling patients most likely to benefit
4. Measure does not assess timing of enrollment and thus amount of exposure to intervention services
5. Measure is based on all enrollees, including those not covered by Medicare or Medicaid

Table III.4. Enrollment effectiveness, by percent change in projections

Percent change between Year 1 and Year 3 participant projections	Number of awardees	Number (percentage) of awardees that met:	
		At least 90 percent of final three-year projections	Less than 65 percent of final three-year projections
Decreased by 10 percent or more	20	13 (65%)	2 (10%)
Stayed within +/- 10 percent	13	4 (31%)	3 (23%)
Increased by 10 percent or more	5	2 (40%)	0 (0%)

Source: Participant projections at the beginning of Year 1 were obtained from first program quarter (September 2014–November 2014) data provided by the implementation and monitoring contractor. Participant projections at the beginning of Year 3 were obtained from 9th program quarter (September 2016–November 2016) data provided by the implementation and monitoring contractor. The number of participants actually served was obtained from 12th program quarter (June 2017–August 2017) data provided by the implementation and monitoring contractor.

The second reason that the findings on enrollment may provide an incomplete picture of enrollment effectiveness is that, in order to reach their original projections, some awardees expanded their program eligibility requirements to include patients who, in the absence of the program, were at lower risk for the inappropriate use of costly medical care. By including these patients, the awardees may have weakened their programs' potential to achieve overall impacts. For example, Johns Hopkins University's program was intended to keep dually eligible beneficiaries with a prior diagnosis of Alzheimer's disease or other dementia-related neurodegenerative diseases in their homes. To boost enrollment, the awardee not only relaxed the program eligibility criteria in the second program year by not requiring beneficiaries to have prior diagnosis of dementia but also allowed Medicare-only beneficiaries to enroll. Johns Hopkins University also lowered its final enrollment projections by half. As a result of these changes, the awardee succeeded in meeting its final (lower) enrollment projection but ended up diluting the potential impact of its program on costs and health outcomes by serving Medicare-only patients who typically have fewer health needs than dual eligible beneficiaries and for whom any savings gained from averted institutionalizations would not accrue to Medicare.

Third, our measure of enrollment effectiveness does not capture the extent to which program participants had the characteristics of the target population and were therefore most likely to benefit from the intervention services in terms of lower costs and improved health outcomes. In most cases, we lacked the clinical information to make this assessment. For example, one of the main components of the University of Michigan's intervention was a newly developed electronic risk assessment tool that surgeons could use to assess their patients' risk of complications following major abdominal surgery. However, many surgeons chose not to use the tool and instead relied on their clinical judgment to identify patients they thought would benefit from the intervention.

Fourth, the findings on enrollment effectiveness do not reflect the timing or duration of program participation, making it difficult to determine when impacts should occur or to evaluate the level of exposure to intervention services. Finally, the number of participants includes all participants regardless of payer. These counts differ (sometimes substantially) from the counts we used to conduct the impact analysis, which is restricted to Medicare (or Medicaid)

beneficiaries who are linked to the Medicare (or Medicaid) enrollment files, meet any claims-based eligibility criteria imposed by the awardee, and satisfy other inclusion criteria we use to define the treatment group.

2. Effectiveness of implementing service delivery models

Two-thirds (25) of the 38 awardees met the criteria for the delivery of services domain, indicating that they succeeded in delivering intervention services consistent with the planned mode, frequency, intensity, duration, and quality (Table III.5). Of these, 17 awardees met the criteria in all three other service delivery domains (staffing, provider engagement, and patient engagement), when applicable. Examples of the awardees' failure to meet the effectiveness criteria for staffing and training, provider engagement, and patient engagement as defined earlier in Table III.5 follow.

Seventeen awardees met the criteria for service delivery implementation effectiveness in all relevant domains.

Domain	Effective	Partly effective	Not applicable
Delivery of services	25	13	
Staffing and training	25	13	
Provider engagement	20	13	5
Patient engagement	26	9	3

- Five of the 25 awardees that delivered services effectively failed to meet the **staffing and training criteria**. For example, Columbia University did not meet the criterion of recruiting and hiring qualified program staff in a timely manner. The awardee encountered significant delays in hiring community health workers (CHWs). And in turn, these delays made it difficult for the awardee to meet the second criterion: providing relevant and effective training in its service delivery protocol. The awardee had to repeat the training several times—each time a new group of CHWs was hired.
- Four of the 25 awardees that delivered services effectively failed to meet the **provider engagement criteria**. Avera, for example, had a goal that SNF staff would contact the program administrators before resident transfers to the ED or to a hospital at least 50 percent of the time. However, the SNF staff contacted the program administrators before resident transfers only 11 to 26 percent of the time per quarter during the cooperative agreement. The lack of program contact by the SNF staff indicates that the awardee did not consistently meet the criterion of engaging participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way.
- Finally, one of the awardees (Nebraska Medical Center) that delivered services effectively struggled to meet the **criteria for engaging participants** in a timely manner and in a meaningful way; the awardee also struggled to retain most participants for the full period of enrollment. Nebraska Medical Center reported that it had trouble with both of these criteria, especially with respect to the sickest and highest-risk patients—the patients for whom the program might have the greatest benefits.

Table III.5. Summary of service delivery implementation effectiveness, by domain

Acronym/ abbreviation	Delivery of program services	Staff hiring, retention, and training	Recruiting and engaging providers and provider organizations	Engaging program participants
AAMC	E	E	E	n.a.
ACCF	PE	PE	PE	E
Altarum	PE	E	PE	n.a.
Amerigroup	PE	E	PE	PE
Avera	E	E	PE	E
BMC	PE	PE	PE	PE
CCC	E	E	E	E
CCNC	E	E	E	E
CHIIC	PE	E	E	E
CHS	PE	PE	PE	PE
Clifford Beers	PE	E	PE	E
Columbia	E	PE	n.a.	E
DMC	E	E	E	E
FPHNY	E	E	E	E
FSCL	E	PE	E	E
Hopkins	E	E	E	E
Icahn	E	E	n.a.	E
Mesa	E	E	E	E
Montefiore	E	PE	E	E
NACHRI	E	E	E	E
NHCHC	PE	PE	E	PE
NMC	E	E	PE	PE
Northwell	E	PE	E	E
NYC H+H	PE	PE	n.a.	PE
SCH	E	E	E	E
U KS	E	E	E	E
U NC	PE	E	PE	PE
UCSD	E	E	E	E
UCSF	E	E	E	E
UHCMC	E	PE	PE	E
UIC	PE	PE	n.a.	PE
UMich	E	E	PE	E
UNM	PE	PE	PE	n.a.
Ventura	E	E	E	E
VillageCare	E	E	E	E
Wash U	E	E	n.a.	E
WI DHS	E	E	E	E
Yale	PE	PE	PE	PE

Source: Review of awardees' self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017; interviews with awardee administrators and frontline staff from selected sites conducted annually toward the end of each of the three years of the cooperative agreements.

Note: A rating of effective (E) means that the awardee met program goals for that domain in a timely manner, in a meaningful and sustained way, and consistent with its ability to improve its targeted outcomes. A rating of partly effective (PE) means that the awardee experienced a delay or other implementation challenge in that domain that threatened its ability to improve its targeted outcomes. An "n.a." indicates that the domain was not applicable to the awardee's program.

The remaining third (13 awardees) failed to meet the criteria for the delivery of services domain (Table III.5). About half of them (6) failed to meet the criteria in the three other domains, if applicable, and four failed to meet the criteria in two additional domains. One awardee (Catholic Health Initiatives) met the criteria for service delivery effectiveness in all three other domains (staffing, provider engagement, and patient engagement) but failed to satisfy the criteria for the delivery of services domain because of differences in how its participating sites defined the intervention. The awardee intended to implement a health coaching intervention targeted to patients with selected chronic conditions. But some of the participating sites deviated from the short-term, intensive care management model of care in ways that were less likely to lead to cost savings and improved health outcomes. Because Catholic Health Initiatives did not have enough staff for routine clinical duties, and because of changes in Medicare covered benefits, the health coaches at some sites began spending more time providing newly billable services such as annual wellness visits, following up with patients who were recently discharged from a hospital, and offering advanced care planning for all patients.

We find that effective implementation of a model is associated with four structural or program design characteristics of the awardees' programs. First, interventions intended to help beneficiaries with (often long-term) chronic conditions (including adults and children with both a low and a high risk of the unnecessary use of medical services) were more likely to have been implemented effectively than interventions for which an awardee recruited participants after a precipitating event (such as a hospital discharge, an ED visit, a fall or an injury, the onset of back pain, or pregnancy). People with longer-term chronic conditions may have a closer relationship with their providers, may not be in immediate pain or discomfort, and may be more willing to try alternative approaches to care than people who recently experienced an unexpected and potentially serious health problem or event. Second, service delivery models that focused on changing provider behavior or transforming the way in which care is delivered (such as using EMR data to manage care or improving the processes of care) were more likely to have been implemented effectively than those focused on changing patient behavior (such as enrolling patients in a new health coaching program or getting them to use new health IT technologies at home). Third, programs serving a single, local market area were more likely to have been implemented effectively than those serving larger or multiple geographic areas. Providers may have been more willing to engage in a program if they knew or had a direct relationship with program administrators and staff. Finally, programs that were based on a pilot program were more likely to have been implemented effectively than programs administered by awardees with little or no experience in providing intervention services to the target population.

Awardees were more likely to have implemented their service delivery models successfully if they:

- Targeted chronic versus acute conditions
- Focused on changing provider versus patient behavior
- Served single and local market areas versus larger and multiple market areas
- Had conducted a pilot test of their model prior to this award

3. Assessment of overall implementation effectiveness

For reasons explained in Section B of this chapter, to be considered implemented effectively overall, we relied on the rating for the service delivery component. To be considered implemented effectively in the service delivery component, an awardee had to meet the criteria in the delivery of services domain (domain #1 in Table III.1) plus (in most cases) the criteria for at least two of the other three domains, when relevant to the program. The one exception to this rule was the University of Michigan. The University of Michigan delivered its services largely as intended at its affiliated practices, but had trouble engaging surgeons who were not affiliated with the University Michigan Health System in the program. The lack of timely and meaningful engagement of surgeons at nonaffiliated practices limited the reach of the program and threatened the awardee's ability to achieve its overall outcome goals. It also has implications for the sustainability of the program, an issue we discuss in Chapter VI.

Based on this methodology, we concluded that 24 of the 38 programs were implemented effectively (see the "Service delivery" column in Table III.6). By this, we mean that the programs were implemented such that we would expect to see lower costs and/or improvements in health outcomes (depending on the goals of the program) if the interventions *as designed* were effective. In other words, the qualitative evidence available to us suggests that, for these 24 awardees, there were no breakdowns in the delivery of services, in staffing and training, or in provider and patient engagement that would threaten the achievement of the programs' intended impacts.

Twenty-four awardees had met the criteria for full implementation effectiveness by August 2017.

- AAMC
- Avera
- CCC
- CCNC
- Columbia
- DMC
- FPHNY
- FSCL
- Icahn
- Johns Hopkins
- Mesa
- Montefiore
- NACHRI
- NMC
- Northwell
- SCH
- UKS
- UCSD
- UCSF
- UHCMC
- Ventura
- VillageCare
- UWash
- WI DHS

Table III.6. Summary of overall implementation effectiveness

Acronym/abbreviation	Enrollment	Service delivery
AAMC	E	E
ACCF	E	PE
Altarum	E	PE
Amerigroup	E	PE
Avera	E	E
BMC	PE	PE
CCC	E	E
CCNC	E	E
CHIIC	PE	PE
CHS	PE	PE
Clifford Beers	PE	PE
Columbia	NE	E
DMC	PE	E
FPHNY	E	E
FSCL	E	E

Table III.6 *(continued)*

Acronym/abbreviation	Enrollment	Service delivery
Hopkins	E	E
Icahn	NE	E
Mesa	PE	E
Montefiore	E	E
NACHRI	E	E
NHCHC	NE	PE
NMC	PE	E
Northwell	E	E
NYC H+H	PE	PE
SCH	PE	E
U KS	E	E
U NC	E	PE
UCSD	E	E
UCSF	E	E
UHCMC	PE	E
UIC	E	PE
UMich	NE	PE
UNM	NE	PE
Ventura	PE	E
VillageCare	PE	E
Wash U	PE	E
WI DHS	E	E
Yale	E	PE

Source: Review of awardees' self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017; interviews with awardee administrators and frontline staff from selected sites conducted annually toward the end of each of the three years of the cooperative agreements.

Notes: For enrollment, a rating of effective (E) means that the awardee met more than 90 percent of its final enrollment target; a rating of not effective (NE) means that the awardee met 65 to 90 percent of its final enrollment target; and a rating of partly effective (PE) means that the awardee met less than 65 percent of its final enrollment target. Awardees were allowed to change their enrollment projections (up or down) at the end of the first and second years of their cooperative agreements. We based our assessment of enrollment effectiveness on awardees' final enrollment count relative to their projections as reported in their ninth quarterly reports.

For service delivery, a rating of effective (E) means that the awardee met program goals for that domain in a timely manner, in a meaningful and sustained way, and consistent with its ability to improve its targeted outcomes. A rating of partly effective (PE) means that the awardee experienced a delay or other implementation challenge that threatened its ability to improve its targeted outcomes.

Of the 24 programs that were implemented effectively overall, the awardees that operated eight of these programs failed to meet the 90 percent enrollment threshold that we used to measure enrollment effectiveness. However, all but one of these awardees succeeded in reaching at least 75 percent of their targets, and six of them enrolled more than 1,300 beneficiaries. Only two awardees (Columbia University and Mount Sinai) failed to meet the 65 percent threshold for partial effectiveness, and only three (Johns Hopkins, Mount Sinai, and Seattle Children's Hospital) enrolled fewer than 1,000 beneficiaries. In all of these cases, we have concluded, based on the qualitative information available to us, that the awardees' success in meeting most or all of their service delivery goals outweighs the shortcoming in meeting their enrollment goals.

Based on our review of the qualitative information, we have determined that the remaining 14 awardees did not reach full implementation effectiveness by August 2017. As a result, we believe that their programs would be unlikely to achieve their intended outcomes even if a fully implemented intervention would have been effective. Six of these awardees met the 90 percent enrollment threshold, but they did not meet the criteria for effectiveness in implementing their service delivery models. Eight awardees failed to meet the criteria for both enrollment and service delivery.

Fourteen awardees had not met the criteria for full implementation effectiveness by August 2017.

- ACCF
- Altarum
- Amerigroup
- BMC
- CHIIC
- CHS
- Clifford Beers
- NHCHC
- NYC H+H
- U NC
- UIC
- UMich
- UNM
- Yale

The next step in our analysis is to use these two awardee groups based on implementation effectiveness (implemented effectively and implemented partly effectively) to identify the factors associated with success, the barriers that impeded progress, and the strategies used to overcome common challenges. Before reviewing these factors, we examine the differences between the two groups in terms of how the awardees' staff and participating clinicians perceived the effects of the programs on the delivery of care and on health outcomes. After we derive the final impact estimates, we will assess the correspondence between implementation effectiveness and patient outcomes.

D. Assessment of perceived program effects on care delivery and health outcomes

Because of the wide variation in target populations and service delivery models, we summarized the survey responses from staff and clinicians' on their perceptions of the programs' effectiveness in four domains that are relevant to nearly all awardees: (1) participant engagement, (2) perceived impact on the quality of care and services, (3) effectiveness in achieving program goals, and (4) staff and clinician burnout. The staff survey was fielded at the end of the second program year, and the clinician survey was fielded about halfway through the third program year. As a result, they represent perceptions throughout most but not all of the program implementation period. Our description of survey results are applicable only to the awardees for which we fielded surveys (35 awardees for the staff survey and 18 awardees for the clinician survey), but these awardees included nearly all awardees for which the surveys were relevant. The surveys included groups of respondents that are not comparable across awardees because of the diverse nature of the programs. As a result, we describe broad patterns in the data overall and as they relate to our assessment of the awardees' overall implementation effectiveness. We do not show numerical results for fewer than 11 respondents because we want to avoid the possibility of identifying individual sample members, but we include their responses in our qualitative descriptions of the data.

Participant engagement. The majority of staff in just over half of the 35 programs that received the staff survey strongly agreed that their program successfully engaged participants (Table III.7). When we expanded the analysis to include additional responses, nearly all staff in nearly all programs "somewhat agreed" or "strongly agreed" that their program had successfully

engaged participants; very few staff disagreed (data not shown). There was little correspondence between the staff survey results on participant engagement and our assessment of overall implementation effectiveness.

Perceived impact on the quality of care and services. The majority of staff in all programs that received the staff survey reported that their programs had a positive impact on the quality of care and services (Table III.7). This was also true for 17 of the awardees that received the clinician survey.[3] There was little correspondence between the staff's or the clinicians' perceptions of the impact of the program on the quality of care and services, and our assessment of overall implementation effectiveness when we compared means between effective and partly effective awardees and assessed the correlation between effectiveness and the proportion of respondents reporting that their program had a positive impact.

[3] Clinicians in the Clifford Beers Guidance Clinic did not receive this survey item due to an error in the survey instrument.

Table III.7. Summary of staff and clinician survey responses on perceived effects of HCIA R2 programs on the delivery of care and outcomes

Acronym/ abbreviation	Overall effectiveness[a]	Engaged participants: strongly agreed	Quality of care and services: positive perceived impact		Achieving program goals: very effective		Burnout: Decreased a little or a lot		Burnout: Increased a little or a lot	
		Staff	Staff	Clinicians	Staff	Clinicians	Staff	Clinicians	Staff	Clinicians
AAMC	E	—	—	88%	—	50%	—	40%	—	9%
ACCF	PE	35%	54%	59%	29%	19%	12%	5%	29%	26%
Altarum	PE	82%	↓	—	73%	—	↓	—	↓	—
Altarum	PE	40%	90%	90%	50%	37%	5%	7%	5%	12%
Amerigroup	PE	67%	100%	—	71%	—	26%	—	13%	—
Avera	E	43%	98%	—	38%	—	21%	—	12%	—
BMC	PE	55%	100%	—	35%	—	10%	—	15%	—
CCC	E	45%	70%	—	29%	—	4%	—	29%	—
CCNC	E	↑	↑	—	↓	—	↓	—	↓	—
CHIIC	PE	48%	84%	85%	54%	22%	22%	17%	30%	30%
CHS	PE	↑	↑	—	↓	—	↓	—	↓[b]	—
Clifford Beers	PE	59%	100%	↑	33%	↑	0%	↓	67%	↓
Columbia	E	75%	88%	↓	50%	7%	26%	0%	13%	21%
DMC	E	55%	95%	—	50%	—	5%	—	35%	—
FPHNY	E	57%	83%	—	57%	—	5%	—	36%	—
FSCL	E	50%	84%	68%	20%	36%	24%	5%	24%	50%
Hopkins	E	62%	76%	—	43%	—	15%	—	34%	—
Icahn	E	↑	↑	88%	↑	88%	↓	33%	↓	29%
Mesa	E	50%	100%	↓	33%	↓	12%	↓	45%	↓
Montefiore	E	61%	91%	81%	48%	34%	17%	43%	48%	18%
NHCHC	PE	↑	↑	—	↑	—	↓	—	↑	—
NMC	E	41%	81%	—	65%	—	12%	—	36%	—
Northwell	E	↑	↑	—	↑	—	↓	—	↓	—
NYC H+H	PE	75%	94%	—	30%	—	15%	—	39%	—
SCH	E	↑	↑	↑	↓	↑	↓	↑	↓	↓
U KS	E	54%	93%	97%	54%	81%	17%	26%	6%	13%

Table III.7 *(continued)*

Acronym/ abbreviation	Overall effectiveness[a]	Engaged participants: strongly agreed	Quality of care and services: positive perceived impact		Achieving program goals: very effective		Burnout: Decreased a little or a lot		Burnout: Increased a little or a lot	
		Staff	Staff	Clinicians	Staff	Clinicians	Staff	Clinicians	Staff	Clinicians
U NC	PE	—	—	63%	—	17%	—	23%	—	16%
UCSD	E	44%	94%	—	44%	—	0%	—	61%	—
UCSF	E	38%	88%	—	25%	—	8%	—	79%	—
UHCMC	E	↓	↑	—	↓	—	↓	—	↑	—
UIC	PE	50%	86%	—	50%	—	6%	—	50%	—
UMich	PE	42%	81%	84%	19%	40%	0%	16%	9%	10%
UNM	PE	24%	74%	↑	15%	↑	9%	↓	15%	↓
Ventura	E	73%	75%	70%	58%	16%	0%	11%	8%	10%
VillageCare	E	↑	↑	—	↑	—	↓	—	↓	—
Wash U	E	77%	92%	↑	42%	↑	8%	↓	38%	↓
WI DHS	E	71%	96%	100%	65%	78%	8%	↓	46%	↑
Yale	PE	55%	77%	—	59%	—	14%	—	18%	—

Source: HCIA R2 staff and clinician surveys, field by Mathematica in July to October 2016 and March to June 2017, respectively.

Note: Cells with a "—" indicate that we did not field a survey with the given respondent type for that awardee. A detailed description of the survey methodology and survey instruments is available from the authors on request. We do not report quantitative results for survey items with fewer than 11 respondents to avoid the potential for identifying sample members. An upward arrow represents 50 percent or more of respondents with the response in the column title. A downward arrow represents fewer than 50 percent of respondents with the response in the column title.

[a] Based on Mathematica's rating of overall implementation effectiveness, where E = effective and PE = partly effective.

Effectiveness in achieving program goals. In contrast to their perceptions of their programs' impact, the staff and clinicians were less positive about their programs' effectiveness in achieving program goals. In both surveys, the staff and clinicians rated the effectiveness of their program in "achieving its goals" on a scale that included very effective, somewhat effective, somewhat ineffective, and very ineffective. The majority of staff in only about one-third of the programs rated those programs as "very effective" in achieving their goals (Table III.7). There was not a clear correspondence between responses to this survey item and our overall rating of implementation effectiveness when we compared means between effective and partly effective awardees and assessed the correlation between effectiveness and the proportion of respondents reporting that their program had been effective in achieving its goals. Similarly, the majority of clinicians in less than half of the programs rated the program as "very effective" in achieving its goals. In 6 of the 10 programs that we rated as effective in overall implementation, the majority of clinicians described their program as "very effective," compared with only 2 of the 8 programs that we rated as partly effective. However, few staff or clinicians rated their programs as "ineffective" or "very ineffective" in reaching their goals. When we analyzed all response options, we found that nearly all staff and clinicians in all surveyed programs rated their program as "very effective" or "somewhat effective" in reaching the program goals (data not shown).

Among the 10 awardees that received both the staff and the clinician survey and that had enough respondents to report the results (7 awardees received both surveys but had too few respondents), there was little correspondence between the two surveys in the respondents' assessments of the program's effectiveness in reaching its goals. The lack of correspondence could be due to measurement error. In addition, the difference in responses between the staff and clinicians could have multiple explanations depending on the structure and context of the awardees' programs, including differences in staff and clinician roles, in the fact that one group may have had a more realistic assessment of effectiveness than the other, and in progress in implementation between the periods when the two surveys were fielded.

Staff and clinician burnout. Staff and clinicians were also asked in the surveys whether the programs increased or decreased their feelings of burnout, based on their own definition of burnout. Relatively few staff reported that the programs *decreased* their feeling of burnout (Table III.7). This finding is not surprising given that reducing staff burnout was not an explicit goal for any of the awardees and that the programs required changes in how care was delivered, which could create additional demands and stress on staff. A higher proportion of clinicians (among the awardees that received the clinician survey) reported that the program decreased their feelings of burnout, and there was a moderate correlation between our rating of implementation effectiveness and the proportion of clinicians that reported their program decreased their feelings of burnout. There were eight programs in which one-quarter or more of clinicians reported that the program decreased their feeling of burnout (the Association of American Medical Colleges, Mount Sinai, Mesa, Montefiore Medical Center, Seattle Children's Hospital, the University of Kansas, Clifford Beers Guidance Clinic, and the University of New

Mexico).[4] More of these awardees (six) were rated effective in overall implementation, compared with two that were rated as partly effective. This is an encouraging finding because it suggests that supports to clinicians from these specific programs show promise for reducing burnout.

However, for more than half of the awardees that received the staff survey, one-quarter or more of the staff reported that the program *increased* their feelings of burnout. This was also true for more than one-third of the awardees that received the clinician survey. This finding suggests that many programs substantially burdened staff and clinicians. There was little correspondence between the staff survey results on increased burnout and our assessment of overall implementation effectiveness. However, there was a trend toward clinicians reporting more burnout among programs rated effective in overall implementation. There were six programs in which one-quarter or more of clinicians reported that the program increased their feeling of burnout (American College of Cardiology Foundation, Mount Sinai, Catholic Health Initiatives, Washington University, Four Seasons, and Wisconsin Department of Health Services), and four of these awardees were rated effective in overall implementation. This is a concerning finding because it suggests that clinician burnout could limit the opportunities for long term success in these innovative programs.

For four of the awardees that received both the staff and clinician surveys—Clifford Beers Guidance Clinic, Four Seasons, Mesa Fire and Medical Department, and Montefiore Medical Center—there was little correspondence between the results of the two surveys with respect to feelings of burnout. Differences in burnout between staff and clinicians could have important implications for the longer-term functioning of the programs. For example, if clinician burnout is reduced by the use of care coordinators who help manage their disease, but staff burnout is increased, then the quality of care could continue to suffer, and the program could face longer-term staffing challenges.

E. Factors associated with implementation effectiveness

1. Factors influencing the awardees' ability to reach their enrollment targets

The factors associated with the awardees' success in reaching their enrollment targets fall into five broad categories: (1) engaging with participating clinicians and external partners in ways that would encourage them to serve as a steady source of referrals, (2) engaging with patients in ways that would encourage them to enroll in the program after a provider referral or through self-referral, (3) building staff capacity and competency in identifying, recruiting, and enrolling patients, (4) developing clear and easy-to-use procedures for identifying and enrolling eligible patients, and (5) obtaining institutional review board (IRB) approval and executing contracts with partner organizations. The awardees that implemented their programs effectively and those that were only partly effective commonly reported that these five facilitators helped them to identify, recruit, and enroll their target populations. The difference in enrollment effectiveness between the two groups has to do with their success in harnessing these facilitators

[4] We do not report quantitative results for survey items with fewer than 11 respondents to avoid the potential for identifying sample members.

to the benefit of their program. We discuss each of these factors below, followed by a review of the unique challenges encountered by the five awardees that failed to reach 65 percent of their enrollment goal.

a. Engaging participating clinicians and external partners to serve as a steady source of referrals

Effective awardees emphasized that building trust with and obtaining buy-in from providers took time and required an ongoing commitment of program resources. Nearly all awardees cited referrals from engaged providers and community partners as critical to achieving their program enrollment goals. In citing this factor, some awardees were referring to the individual physicians internal to the implementing sites whom the awardees expected to identify, recruit, and enroll their patients into newly created services lines, such as health coaching or care coordination. Other awardees were referring to the external partners, including clinical and non-clinical providers of support services. The awardees expected the partners to inform their clients about the program and to encourage them to enroll if the partners believed that the clients would benefit from the program.

Awardees that met their enrollment goals emphasized that building trust with and obtaining buy-in from providers took time and required an ongoing commitment of program resources. Strategies included the following:

- Offering clinicians financial incentives, free medical devices, and continuing medical education credits for participating
- Leveraging their existing relationships with provider organizations
- Using physician champions to build trust and maintain momentum among their colleagues
- Holding regular meetings with and making on-call resources available to providers to answer their questions
- Motivating providers by giving them monitoring reports and other educational materials that would help providers to understand the program
- Embedding program staff in a physician's office or organization
- Creating community outreach events and information campaigns to promote the program

b. Engaging with patients in ways that would encourage them to enroll in the program after a provider referral or through self-referral

As they did with providers, effective awardees stressed that engaging patients, gaining their trust, and convincing them to enroll required a concerted effort and an investment of time and resources. They also said that the process was self-reinforcing: successfully engaging patients early on made it easier to enroll others later. One-third of awardees reported that effectively engaging with eligible patients after they were identified was an important component of enrollment success. Engaging typically meant using culturally and linguistically appropriate messaging, delivered in person and through public outreach, to educate patients about the benefits of the program. For example, Washington University used patient and community advisors as well as marketing and media consultants to develop outreach strategies and messaging to raise awareness of the program. Washington University also reported that word of mouth within the patient community about the clinic's accessibility and services encouraged self-

referrals. Johns Hopkins also intensified its outreach and recruitment efforts through direct mail campaigns. Boston Medical Center recruited families of children with complex needs during face-to-face meetings in their homes or at their medical appointments. Several awardees offered material inducements to enroll or to more fully engage in the program. Ventura provided nicotine replacement therapy to at-risk COPD patients to increase their engagement in the CATCH program, and Village Care provided a $35 per month incentive to participants.

c. Building staff capacity and competency in identifying, recruiting, and enrolling patients

Roughly one-half of the awardees that cited building staffing capacity and competency as a facilitator emphasized the importance of using dedicated staff, often embedded in the referring provider's site or organization, to recruit and enroll patients. The awardees also stressed the importance of training staff to identify (often through review of medical records or other administrative data) potentially eligible participants and to effectively engage with them about the program. Five of the awardees that fully or partly met their enrollment targets reported that they used program funds to hire new staff or to retrain existing staff to identify and enroll patients. The University of California at San Francisco hired bilingual staff to assist in recruiting monolingual populations. Yale University hired translators and multilingual staff to increase enrollment in non-English-speaking communities. Montefiore Medical Center overcame initial lags in enrollment by retraining its staff to screen patients for program eligibility. National Healthcare for the Homeless reported that training in motivational interviewing enabled its staff to successfully approach a patient population that might be reluctant to trust other people. Clifford Beers Guidance Clinic reported that a nurse embedded at Yale New Haven Hospital helped to increase enrollment in the last two program years through in-depth knowledge of the program's eligibility criteria, direct access to the hospital's EMR system, and understanding of physicians' workflows. These factors made it easier to identify potentially eligible children and to orchestrate warm handoffs between the hospital clinicians, the program care coordinators, and the nurse.

Several awardees that failed to meet their enrollment projections stated that lack of capacity or knowledge to identify and enroll patients limited their success. This was particularly true when direct care service providers were used to recruit and enroll patients. For example, Altarum reported having difficulty identifying staff who lived in remote and rural service regions to recruit and train medical providers in those areas. Delays in program staffing and training delayed recruitment and enrollment in the National Health Care for the Homeless Council program as well.

d. Developing clear and easy-to-use procedures to identify and enroll eligible patients

About half of the awardees emphasized the importance of having streamlined and effective processes for identifying eligible patients that minimized the burden of getting patients into the program for both providers and patients. The most commonly cited (and challenging) aspect of the patient recruitment and enrollment process was the effective use of electronic data (such as medical records, population health reports, disease registries, and claims) to identify eligible patients. Ten awardees cited challenges using EMR data to identify patients as a significant barrier to meeting their enrollment targets. These challenges included difficulty using new EMR

Awardees that met their enrollment goals used a range of strategies to overcome challenges using EMR data to identify eligible program participants, including the following:

- Embedding staff at locations where patients' data were stored
- Working only with providers willing and able to obtain and effectively use the data
- Building on existing systems rather than creating new ones
- Refining the methods of identifying eligible patients to fit the data that the awardees had

systems, problems linking records from multiple sources, lack of information to identify disease stage, difficulty building population health or disease registries, outdated patient contact information, and institutional requirements in contracting with local managed care plans to obtain data. In addition to the effective use of EMR data to identify eligible patients, several effective awardees emphasized the importance of in-person handoffs between referring providers and intervention staff to reduce patient anxiety and ensure that the connection was made.

e. Obtaining IRB approval and executing contracts with partner organizations

Six awardees reported that the lengthy process involved in securing IRB approval and patient consent delayed enrollment at some sites. The American College of Cardiology Foundation, for example, reported experiencing continuous challenges with obtaining patient consent due to duplicative paperwork and patients' skepticism about sharing personal data. Local IRB review and consent processes created enrollment delays at some of the National Association of Children's Hospitals' sites as well. Five other awardees reported that signing participation agreements with independent clinics and setting up data use agreements with managed care plans to identify eligible patients took longer than expected and also delayed enrollment. For example, the University of Michigan said that long processes to secure data use agreements with some of its implementing sites created significant delays in enrollment. The bureaucracy challenges involved in partnering with facilities outside the Ventura County Health Care Agency system stymied that awardee's enrollment efforts as well. The National Association of Children's Hospitals found it easier to recruit and contract with affiliated practices than with independent primary care practices.

In addition to difficulties implementing the effective enrollment strategies outlined above, the five awardees that failed to reach at least 65 percent of their enrollment goal targeted patients who were particularly challenging to recruit. These awardees also used active enrollment, so they required patients' (or parents' or caregivers') consent and offered an opt-out option. In some cases, awardees struggled to find and engage these patients given the lack of contact information and competing concerns.

The five awardees that failed to reach 65 percent of their enrollment goal targeted patients who were particularly challenging to recruit. They targeted beneficiaries who:

- Were difficult to reach, such as homeless people and Medicaid beneficiaries (Columbia and NHCHC)
- Had competing health and social service needs (NHCHC)
- Had recently experienced a major medical event or were scheduled to have surgery (UMich and UNM)
- Met the medical requirements for respite or hospital care (Icahn and NHCHC)

In other cases, it was difficult to determine eligibility. Mount Sinai reported that it had trouble determining whether patients met some of its eligibility criteria, such as living at home, having a safe home environment with access to a single (not shared) bathroom, and having enough support for or the ability to complete necessary daily activities. The experience of these awardees suggests that enrolling unstable and vulnerable populations, and those facing immediate competing health and social needs is inherently more difficult than enrolling traditional Medicare beneficiaries with chronic conditions.

2. Factors influencing the implementation of service delivery models

Most effective awardees shared several factors that facilitated service delivery, and the absence of the same factors were barriers to service delivery for many of the partly effective awardees. Many of the factors that influenced successful enrollment also influenced the effective implementation of service delivery models. The most common categories of factors that influenced the implementation of the service delivery models are summarized below. The summary includes selected examples of facilitators as well as examples of how the absence of these factors became barriers for some of the partly effective awardees (or how the presence of some external factors became barriers).

Six factors contributed to awardees effectively implementing their service delivery models:

1. Engaging participants by tailoring services to their needs
2. Engaging providers and community organizations through education and understanding their needs and constraints
3. Creating a supportive internal environment through strong support structures and program champions
4. Hiring strong staff, and training and mentoring them
5. Deploying health IT that helps to meet the needs of participants and providers
6. Having state, local, or institutional regulations and policies that are consistent with program goals, and an adequate local supply of health care resources, providers, and labor

a. Engaging participants by tailoring program services

The awardees kept participants engaged in their programs by tailoring services to their needs, building rapport with them, furnishing services perceived as valuable by them, and offering financial incentives to them. Nearly half of the effective awardees explicitly named participant engagement as a facilitator of service. For example, the health coaches in Nebraska Medicine's RIISC program were able to delve into patients' motivation and the challenges they faced in order to work with them to improve their diabetes self-management. The health coaches were also able to intensify their involvement with patients who needed more monitoring and follow-up. The health educators in the University of Kansas' transitional and chronic care management arms emphasized the establishment of strong relationships and rapport with patients.

In contrast, nearly two-fifths of the partly effective awardees found that the inability to keep participants engaged in the program was a barrier to service delivery. For instance, participants in Yale's PRIDE program were resistant to the home visits that were part of the program. Similarly, it was difficult to contact participants in New York City Health + Hospitals' ED Care

Management Initiative because they either did not trust the program enough to provide accurate contact information, or they did not remember that they were enrolled in the program.

b. Engaging providers and community organizations

This facilitator of service delivery effectiveness included education and marketing to providers and organizations, providing financial incentives, understanding the constraints and needs of providers and organizations, and furnishing services perceived as valuable. Almost half of the effective awardees reported that engaging providers and community organizations facilitated the delivery of services. For example, Four Seasons' efforts to educate the public and local providers about its CPC program helped to dispel misconceptions and mistrust of palliative care. The reputation of Northwell's leaders of the Healthy Transitions program among providers who treat chronic kidney disease helped in the marketing of the program. Because the Cleveland Medical Center's nurse care coordinators and palliative care physicians communicated effectively, the cancer disease teams at Seidman Cancer Center were able to see how the LINCC program could make patient care easier for them by addressing patients' physical, social, and emotional needs.

However, for nearly three-quarters of the partly effective awardees, the failure to fully engage providers formed a barrier to service delivery. The primary care providers in University of North Carolina's BBC program did not fully use the program's checklist because (1) they faced more pressing patient needs, (2) they were dealing with numerous other quality metrics, and (3) the checklists were never fully integrated into the EMRs of several of the practice sites. Likewise, the dental practitioners in Altarum's MCPP did not have much experience with performance measurement and technology; this created a barrier to the adoption of the program's MiDR health IT tool. According to program staff, performance measurement is, in fact, not yet part of dental practitioners' culture.

c. Creating a supportive internal environment

This facilitator of service delivery included a team-oriented culture, strong communication and supportive structures within the program, strong support of the program from the administration of the awardees' and the programs' home organization, and the presence of champions for the program. Three-fifths of effective awardees' service delivery models benefited from a supportive internal environment. Program champions in CareChoices' PCCC program encouraged multidisciplinary teamwork and continuous performance improvement, and program leadership successfully fostered communication and collaboration among SNF staff. Program leaders and staff of the Wisconsin Department Health Services' SNP program reported that there was strong support for the program within the department and in the two partner hospitals. It worth noting that nearly half of the partly effective awardees also cited support from leaders, teamwork, camaraderie, and strong communication as facilitators of their service delivery models, and only one partly effective awardee (less than one in ten) complained that their internal environment lacked these elements, so in their minds, their internal environment was not a barrier to service delivery.

d. Hiring, training, and mentoring strong staff

Awardees attended to staffing issues by hiring or deploying enough staff with the appropriate skills for the program, mentoring program staff, providing high quality training for program staff, and achieving high staff satisfaction and therefore staff retention and stability. Slightly over two-thirds of the effective awardees cited one or more of these sorts of attention to staffing issues as an important facilitator of service delivery. The staff of Detroit Medical Center's Gateway program, for example, reported high levels of staff satisfaction and experienced very little turnover. In its INSPIRE program, the Fund for Public Health in New York implemented a telementoring program to teach providers how to treat hepatitis C. This type of instruction was not readily available elsewhere or before the program was launched. Yet, almost three-fifths of the effective awardees still reported that staff turnover and difficulty achieving adequate staffing were barriers to effective service delivery.

Over half of the *partly effective* programs cited staffing issues as barriers to service delivery. The staff of Boston Medical Center's 4C program noted that they (1) had difficulty maintaining the scope of services because of an increase in their caseloads, (2) did not have enough formal training or written protocols, and (3) struggled with the enormity of the medical and psychosocial challenges faced by children with medical complexity. The staff of New York City Health + Hospital's ED Care Management Initiative felt so pressured to recruit and enroll new participants that they did not have enough time to provide comprehensive care management; nor did they feel that they had enough resources to identify and address all the social challenges to health that their participants faced. The staff of Catholic Health Initiatives felt that the geographic expanse of their program's service area and the large number of sites was a barrier to the introduction and implementation of program services. The awardee's health coaches also felt that, because they were in resource-constrained rural clinics, the many other demands on their time limited the amount of time they could spend on the project. Staff of the University of Michigan's MSHOP felt that the university's hiring regulations were a barrier to hiring new staff for their program's coordinating center. The regulations included requirements that all open positions funded by a grant or a cooperative agreement-funded positions must be posted as term limited, and that all open positions in general must be posted for a specified period of time.

e. Deploying health IT that meets the needs of participants and providers

Slightly over one-third of the effective awardees cited deployment of health IT that met various specific needs as a facilitator of service delivery. The Association of American Medical Colleges' ability to integrate its specialty templates into the practices' EMRs facilitated communications between primary care physicians and specialists. Four Seasons' tele-health component enabled it to reach rural patients. VillageCare's appealing and easy-to-use mobile platform, Rango.net, facilitated the implementation of its Rango program, which was focused on improving the adherence of people who are HIV-positive to their antiretroviral medications. Montefiore Medical Center's BHIP program introduced technological tools that increased communication between patients and providers, improved patient engagement, and helped providers to work more efficiently.

One-third of the partly effective awardees found health IT to be a barrier to service delivery. Catholic Health Initiatives' disease registry, which the awardee had hoped would be used by

health coaches to identify participants, was not ready for use until the third program year, and different health IT systems throughout the awardee's network limited the exchange of data between the sites and the ACO; it also interfered with the awardee's effort to engage the sites and improve quality. In the first year of its program, Yale University encountered problems with the quality of the participant data entered into its REDCap IT tool, which were an initial barrier to program implementation. (These problems were later remedied with quality assurance procedures from the second year onward). The American College of Cardiology found it difficult to integrate the tools for its SMARTCare program into its sites' EMRs; this was a constant challenge throughout the duration of the cooperative agreement, and some of the sites never fully integrated all the tools. In each of our annual site visits to the University of Illinois's CHECK program, the staff complained about the poorly functioning care coordination software. This made it necessary for the staff to find work-arounds, which were, in turn, a barrier to program implementation.

f. Having external circumstances that are favorable to the program

External circumstances include various state, local, or institutional regulations and policies, and a supply of health care resources, providers, and labor. This factor facilitated almost one-fifth of the effective awardees' service delivery models. Washington University's C3 program had, and continues to have, funding from the federal Title X family planning program. This funding allowed the program to provide services to uninsured participants at little or no cost and to obtain discounted drugs through the 340b program, which enabled the awardee to keep long-acting reversible contraceptives in stock for same-day insertion. Furthermore, support from Washington University's billing staff and infrastructure allowed program staff to deal with insurance requirements.

Medicare's value-based purchasing program for SNFs encouraged facilities in Avera's program to use the eLTC services to reduce hospital readmissions. For the Mesa Fire and Medical Department, local policy changes that (1) permitted community medicine units to transport low-acuity patients directly to the hospital for additional testing, rather than dispatching a third-party ambulance; and (2) improved care for participants and care coordination with the ED facilitated the implementation of the awardee's CCRI. Finally, the SNFs participating in CareChoice Cooperative's HCIA R2–funded PCCC program were also participating in other complementary grant-funded campaigns and quality initiatives, which facilitated the implementation of the HCIA R2–funded program. Yet slightly over one-third of the effective awardees found external factors to be barriers.

One-third of the *partly effective* awardees encountered external circumstances that were *barriers* to service delivery. Some barriers were within the programs' parent institutions. For example, bureaucratic hiring processes, burdensome documentation procedures, and challenging data systems within the larger New York City Health + Hospitals organization slowed implementation of the ED Care Management Initiative. Similarly, the University of Illinois' institutional policies and requirements related to executing contracts, formalizing partnerships, and hiring and retaining program staff impeded the implementation of the CHECK program. Other barriers were external; for example, New York City Health + Hospitals encountered a local competitive job market, and the Children's Home Society encountered logistical issues in

offering a school-based intervention. National Health Care for the Homeless found that there were a limited number of independent medical providers who could prescribe first-line pharmacotherapy for smoking cessation and influenza vaccinations. Altarum's program was hindered by organizational changes at the Michigan Department of Health and Human Services, which slowed the development and deployment of the MiDR tool. Staff turnover in some hospitals impeded the University of New Mexico's Access program, which led to the loss of institutional knowledge and the need to re-engage providers. Only 2 of the 15 partly effective awardees (a little over 10 percent) encountered external circumstances that were facilitators to implementing their programs.

In sum, the six factors described above were identified as either a barrier or facilitator by at least half the awardees.

- Of these six factors, the most frequently cited (by 36 of the 38 awardees) was attending to staffing issues. Staffing issues were cited by 14 awardees as a facilitator and by 22 awardees as a barrier. Programs that were successfully implemented and those that were not were fairly evenly divided on whether they experienced staffing issues as a facilitator or a barrier.
- Four other factors (participant engagement, provider engagement, supportive internal environment, and health IT) were cited by 24 to 28 of the awardees. For all but one of these factors, the pattern was similar to that described above. Each factor was cited by a majority of both successful and unsuccessful awardees. But the awardees in both groups were fairly evenly divided on whether the factor was a facilitator of their success or a barrier that made it difficult to achieve their service delivery goals.
- The exception to the above pattern was a strong internal environment. This factor was cited as a facilitator by 14 of the 24 successfully implemented programs, but as a barrier by only 2 of them. Similarly, it was cited as a facilitator by 7 of the 15 partly successful programs, but as a barrier by only one of them. Thus, half or more of the awardees viewed a strong internal environment as a clear facilitator, but very few viewed the lack of such an environment to be a detriment to their success.

Three other factors emerged as facilitators of service delivery for a few awardees. They included (1) using processes and protocols to improve and standardize the way in which care is delivered (for example, collecting and acting on self-monitoring data or instituting in-person handoffs from physicians to care coordinators); (2) being flexible in the types and timing of services provided depending on the needs of providers or participants; and (3) having relevant experience in implementing a similar program or in serving the target population.

F. Implication for the design and interpretation of the impact evaluations

The implementation issues discussed above have various implications for the evaluability of the programs and how we design the impact analysis. These implications go beyond whether the program was implemented effectively. Changes in targeting and intensity and the timing of full implementation are also important for the impact evaluation. In order to respond to these implications, we must carefully integrate the implementation findings into the impact evaluations as explained below.

- **Expanding eligibility criteria to patients who are at low risk of the unnecessary use of costly medical care will weaken overall program impacts.** Roughly a quarter of all awardees (particularly those who had difficulty reaching their enrollment targets) expanded their eligibility criteria to make it easier to reach their goals. The patients who became eligible under the expanded criteria tended to be less at risk of adverse outcomes and may have required different types or intensity of intervention. Thus, it will be necessary to evaluate impacts separately by eligibility subgroups defined by risk, if possible.
- **Delays in implementation may reduce participants' exposure to intervention services and influence the timing of anticipated impacts.** Even if an awardee was assessed as fully implemented, that implementation might not have reached maturity until fairly late in the evaluation period. Using information from the implementation analysis on when that point was reached, we will specify hypotheses in advance for each awardee as to when we might reasonably expect to first see impacts emerging.
- **Changes in intervention design during the cooperative agreement may also affect the timing of anticipated impacts.** In addition to delays in implementation, some awardees also modified their intervention during implementation and were judged to have successfully implemented that modified version. Such changes were fully expected for innovative programs, but they could create problems for interpretation of the findings and our ability to detect impacts. If all observations are included in a single evaluation sample, the lack of effects encountered during the early period before the design change will lead to underestimation of the effects of the model after the design change. Thus, we will test for whether program impacts changed after major changes in the intervention. However, observed differences in effects could also be a result of program maturation. Ideally, we will be able to identify some outcomes that are more likely than others to be affected by the change in the intervention, or to differences across sites in when the intervention changes were implemented. These hypothesized differences can then be tested in our analysis.
- **The absence of formal intervention protocols or a deviation from the theory of action/theory of change will make it difficult to define the intervention.** Some awardees had well-developed operational protocols derived from their theories of change, while others did not. This is a typical problem in evaluating innovative and evolving interventions. Even if the program is found to produce savings or improve health outcomes, it cannot be effectively expanded or replicated without a clear description of precisely what comprised the intervention and how it was tailored to individual participants. Based on our understanding from interviews with awardees, we will describe the implemented intervention and how it deviated from what was planned and why. We will also describe the version of the program that is being replicated or scaled for awardees that are making such efforts.
- **Similar pilot programs implemented before the launch of the HCIA R2 program will make it difficult to define a pre-period.** If the awardee was implementing the intervention before HCIA R2 and our sample members were included in that prior testing, our difference-in-differences approach may lead to estimates of little or no impacts because the treatment group may exhibit relatively small changes in outcomes. In some cases, we may be able to go back slightly further in time to define a true "pre-intervention" baseline period. In other

cases, it may be necessary to select a comparison group that matches our treatment group in the pre-program period on beneficiary characteristics and outcomes that would not have been affected for the treatment group in the pre-period. We would then simply compare treatment and comparison group outcomes in the follow-up period.

- **Failure to meet enrollment targets may reduce power and make it difficult to observe program impacts.** Although power will be reduced if enrollment is lower than planned, it may still be sufficient to detect moderate effects. Thus, an enrollee's failure to reach its enrollment target does not make it infeasible to detect a meaningful effect. Conversely, success in meeting a low target may still yield a sample size for analysis that is too small to detect even large effects. We will use Bayesian analysis and other methods to reduce the variance of our estimates and provide estimates of the probability that effects exceed some minimal level needed for policy relevance. By reducing the dependence on formal hypothesis tests to draw our inferences about program effectiveness, statistical power becomes less relevant. Nonetheless, the precision of the estimates remains highly important, so we will take numerous steps to offset the deleterious effect of smaller-than-expected sample sizes. These steps include looking across time periods and related outcomes in drawing conclusions about true effects, trimming outliers, and looking for consistency between our impact estimates and our findings on implementation effectiveness.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE IMPACT EVALUATION

A. Introduction and summary of findings

We have conducted a detailed reassessment of the evaluability of each of the 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. **Section B** provides an overview of the quasi-experimental impact evaluation design we use when a randomized control trial was not possible (only two awardees have randomized controlled trials). Appendix A presents the key evaluation challenges that remain with respect to the identification of credible comparison groups—the most significant challenge in the use of quasi-experimental designs for program evaluation. The appendix also explains our proposed strategies for addressing these challenges and describes the quality of the matched comparison groups developed for four awardees.

Section C describes the criteria we used to reassess "evaluability"—the level of rigor possible in an impact evaluation—for each awardee. We then identify the awardees for which we believe we can conduct a rigorous impact analysis and summarize the findings within each of six mutually exclusive awardee groups defined on the basis of the health problems and the age of the awardee's primary target population. For awardees for which we do not believe we can conduct a rigorous impact analysis, we summarize the reasons for this decision, and we describe the data analysis that we plan to conduct for these awardees.

The final two sections present our actual data analysis to date. **Section D** describes baseline characteristics of the Medicare beneficiaries in 27 programs and of the Medicaid beneficiaries in 5 programs. **Section E** provides a synthesis of our preliminary impact findings for three awardees.

The key findings from the implementation evaluation include the following:

- We expect to produce rigorous beneficiary-level impact estimates for 23 awardees (9 for Medicare only, 8 for Medicaid only, and 6 for both Medicare and Medicaid). Well-matched comparison groups have been selected for 3 awardees to date, and 2 others have control groups arising from a randomized controlled trial.
- The early impact findings for 3 of the 23 awardees (shown in Table ES.1) are mixed and should be interpreted with caution because (1) the analysis included only beneficiaries enrolled in a program during roughly the first two years of the cooperative agreement and (2) outcomes were measured during only the first year after enrollment. Estimated impacts on total expenditures were not statistically significant for any of the 3 awardees. For 1 of the 3 awardees (the University of Illinois at Chicago, or UIC), statistically significant estimates of program impacts were favorable for several outcomes. UIC was not among those awardees that we assessed as having effectively implemented its program.
- We will not be able to produce impact estimates for the remaining 15 awardees because they have too few enrollees, their eligibility criteria cannot be replicated for a comparison group,

and/or their primary outcomes are not obtainable from claims data. For these awardees, we will conduct alternative analyses, such as a comparison of outcomes among treatment beneficiaries in the pre- and post-award periods.

B. Overview of optimal impact evaluation design

Because randomized trials (generally considered the gold standard for evaluations) are not typically feasible, we will be using propensity score models to select comparison groups that have characteristics that closely resemble those of the treatment groups. We will then estimate program effects by using a difference-in-differences model whenever possible. The model rests on a particular assumption: that in the absence of the intervention, the change in outcomes over a specified period would have been the same, on average, for the treatment group and for the comparison group. Under this "parallel trends" assumption, any treatment-comparison difference in the change in mean outcomes must be the result of the intervention. We therefore match treatment beneficiaries to potential comparison beneficiaries on measureable characteristics that may influence study outcomes. Although treatment and comparison groups should be well balanced (that is, have similar characteristics on average) after matching, we will use regression adjustment to control for observable characteristics of the study sample to ensure that the estimates are as robust as possible. The regression model will also yield somewhat more precise estimates of program impacts.

Most of our impact analyses will use an intent-to-treat design. That is, sample beneficiaries will be included in the treatment group from the time of enrollment or participation in a program until the end of the evaluation period, even if the beneficiary withdraws from the program or no longer receives intervention services from the awardee. Similarly, we will follow matched beneficiaries in the comparison group from the time of "pseudo enrollment"—or from the time they were attributed to a comparison provider—until the end of the evaluation period.[5]

Where it is feasible, we will use panel data for estimation—that is, we will compare the change in outcomes between baseline and follow-up for program participants to changes over the same period for members of the matched comparison group. For a few awardees, the estimates will be based on different individuals in the baseline and follow-up periods who are served by the treatment and comparison group providers in the two periods. This approach will be needed, for example, when participants are not always observed in the period before program entry, as is the case for children with complex medical needs in the programs of the National Association of Children's Hospitals and the Wisconsin Department of Health Services.

We will use claims data to construct outcome measures for each beneficiary in both the treatment and comparison groups during the intervention period for up to three years. The baseline year will be the 365 days preceding enrollment in the program when we use panel (longitudinal) data. When different people comprise the pre- and post-intervention samples, the

[5] In some cases, pseudo-enrollment dates are assigned to comparison group beneficiaries to correspond to program enrollment dates for treatment group beneficiaries. When the intervention for an individual begins with a triggering event, such as a hospital discharge or entry into a nursing facility, that event will define the pseudo-enrollment date for the comparison group.

pre-intervention period will ordinarily be defined as the 365 days before the program launch date. More detailed information on our methodological approach is provided in Appendix B.

C. Summary of program evaluability

This summary of the awardees' evaluability is derived from information from the evaluability assessments, which we have updated based on data on program implementation and enrollment up to August 31, 2017. The summary begins with an overview of the criteria we used to reassess evaluability to determine the level of rigor in the impact evaluation and concludes with the findings from our reassessment. Our conclusions about evaluability are essentially unchanged since the Second Annual report. For a few awardees, however, we note the additional information we are gathering that may change that assessment.

1. Summary of the criteria used to assess program impacts

We applied three criteria, described below, to all HCIA R2 awardees to assess whether each of the programs could be rigorously evaluated:

Is the sample size sufficient to detect a 10 percent or 20 percent effect? Using actual enrollment as of August 31, 2017, based on files submitted by the awardees, we calculated the cumulative number of Medicare fee-for-service (FFS) and Medicaid FFS and managed care beneficiaries who were enrolled in each program that ended enrollment at that time. For those awardees continuing to enroll new beneficiaries, we projected the number of enrollees who will be enrolled and will have some minimum level of exposure to the program as of February 28, 2018.

We used information on the distribution of expenditures in the baseline period to calculate whether each program's total enrollment was sufficient to detect, with 80 percent power, a program effect of 20 percent or greater on mean expenditures using a two-tailed test at the 0.10 significance level.[6] Effects larger than 20 percent are relatively unlikely for interventions such as the ones being provided by HCIA R2 awardees. For 30 awardees, we were able to calculate sample size requirements based on the variance and mean of key outcome measures we calculated from baseline data for beneficiaries enrolled in their programs before August 31, 2017.[7]

Can the key outcomes that the program is expected to affect be measured by using Medicare or Medicaid claims? We will use administrative (that is, claims) data to measure core outcomes and some other key enrollee-level outcomes for both the treatment and comparison groups. The four core outcomes are (1) total Medicare or Medicaid expenditures, (2) number of all-cause hospitalizations, (3) number of ED visits (and observation stays) that do not lead to a

[6] That is, we calculated the sample size needed to be 80 percent confident that if the true (unknown) program effect on expenditures were 20 percent of the mean or larger, we would correctly reject the hypothesis that the program had no effect.

[7] For the remaining awardees, we continued to use expenditure and hospitalization statistics (that is, mean and variance for individuals) from published studies of similar populations.

hospitalization, and (4) 30-day unplanned hospital readmissions. For some awardees, the primary expected outcomes may not be fully observable in administrative data or within the model testing period. For example, the University of California at San Diego's program focuses on reducing the 10-year risk of stroke and heart attack. Measuring the reduction in risk requires clinical data, which will be available for the treatment group but not for a comparison group.

Can we identify a credible comparison group by using claims data? There are two primary issues when constructing a credible comparison group for quasi-experimental designs: (1) concern about participation or selection bias and (2) the ability to identify the treatment and comparison groups by using claims data. In order to construct a credible comparison group, we need to select beneficiaries whose characteristics (as measured in claims data, such as diagnoses or types and amounts of service use) are well-matched to those of the treatment group. If participation depends upon unobservable characteristics, we need to consider whether there is any way to approximate the unobserved characteristics with observable data (for example, through the use of proxy variables, such as the characteristics of a patient's zip code, to represent socioeconomic characteristics). Alternatively, in some cases we redefine the treatment and comparison groups so that both groups are defined in exactly the same way, using only observable characteristics (for example, evaluating program effects on all those who met the program's claims-based eligibility criteria rather than only on those who actually received the intervention). Obtaining credible impact estimates is challenging when (1) beneficiaries have the option of whether to participate in the program, or when programs selectively recruit beneficiaries, and (2) the participation rate among all beneficiaries eligible for the program is low. Therefore, we do not typically consider an intervention to be evaluable if the participation rate among eligible beneficiaries identifiable in claims data is below 20 percent.

Except for those awardees implementing a randomized design and providing us with the randomized control group, we will draw a comparison group by using Medicare and Medicaid claims data as our primary data source. A number of awardees identify beneficiaries for the treatment group based on the occurrence of an acute event, such as an ED visit for an emerging stroke, or a pattern of high expenditures or excessive acute care utilization. In these instances, claims data are a good source of information for identifying both the treatment and the comparison group within an intent-to-treat evaluation framework. Other awardees, however, rely on the clinical judgment of providers or on non-claims data to identify the treatment group; these data are seldom available for a potential comparison group. Absence of comparable data from which to draw a comparison group will require us to evaluate program effects on a group of individuals in a redefined treatment group and an analogous comparison group, both of which we can identify through the same algorithm that draws on information from Medicare or Medicaid enrollment and claims. This process will result in a treatment group that includes some eligible individuals who were never enrolled and may exclude some individuals who were enrolled but did not meet the claims-based criteria we use. The matched comparison group will be identified by using the same claims-based algorithm. If the algorithm results in the inclusion in the treatment group of some people who were never enrolled, then the estimated treatment effect will be diluted in this intent-to-treat framework.

For each awardee, we will select a comparison group tailored to the particular context of the intervention, including markets, providers, and beneficiaries involved. This criterion reflects the assessment of our ability to draw a solid comparison group with claims data that reflect the treatment group (or some subset or tightly defined superset) on observable characteristics. For example, it would be extremely difficult to identify a comparison group for an awardee that targets homeless Medicaid beneficiaries or children at risk of developing dental problems.

2. Summary of evaluability assessments

Using the three criteria described above, we expect to conduct a rigorous impact analysis for 23 of the 38 awardees (Table IV.1). For the other 15 awardees, we note that we have "serious concerns" about being able to construct a credible comparison group, typically because of a high likelihood of severe selection bias. The discussion below describes the nature of our concern for each of these awardees. Across the 23 awardees, we will conduct nine evaluations that include only Medicare FFS beneficiaries and eight evaluations that include only Medicaid beneficiaries. For 6 of the 23 awardees, we will evaluate program impacts on both Medicare FFS and Medicaid beneficiaries. Detailed assessments are provided in the individual Awardee Program Narratives portion of this report.

It is important to note that our current assessment of evaluability is based on assumptions that may change for a number of reasons: (1) the level of enrollment for any awardee continuing to enroll patients through February 2018 may differ from our projections based on the awardee's prior enrollment rate; (2) the variability in outcome measures during the intervention period may be greater or lower than it was during the baseline period, and as a result the necessary sample size may be greater or lower than we have calculated here; (3) Medicaid data may not be available soon enough to conduct analyses for the final evaluation report; or (4) we are not able to draw a credible comparison group because we cannot replicate the awardee's stated recruitment criteria or because claims data do not support a comparison of beneficiaries eligible for the treatment group to comparisons who are eligible but outside of the awardee's catchment area. Although there are a number of awardees for which a rigorous impact analysis is not likely, we will continue to monitor their enrollment numbers, the characteristics of their participants, the potential for selection bias, and the availability of data for our planned analyses.

Table IV.1. Evaluability assessment for HCIA R2 awardees

Awardee group	Number of awardees	Rigorous impact analysis likely for Medicare beneficiaries	Rigorous impact analysis likely for Medicaid beneficiaries	Rigorous impact analysis not likely
Youth with complex conditions	5		NACHRI, SCH, WI DHS, UIC	BMC
High-risk chronic conditions	8	NM, NYC H+H, UHCMC, UKS	NYC H+H	DMC, FSCL, Northwell, UCSD,
Lower-risk chronic conditions	6	ACCF, CCNC, CHIIC, FPHNY, Ventura, VillageCare,	CCNC, FPHNY, Ventura, VillageCare	
Behavioral health and cognitive disorders	5	UCSF	Clifford Beers, Montefiore	Amerigroup[a] Hopkins,[b]

Table IV.1 *(continued)*

Awardee group	Number of awardees	Rigorous impact analysis likely for Medicare beneficiaries	Rigorous impact analysis likely for Medicaid beneficiaries	Rigorous impact analysis not likely
Acute and sub-acute conditions	6	Mesa	Mesa	Icahn, NHCHC, UMich, UNC, UNM[c]
Primary and preventive care	8	AAMC, Avera, CCC	Altarum, Columbia	CHS,[d] Wash U, Yale[e]

[a]Amerigroup has substantially increased its enrollment, but we continue to have concerns that the primary outcomes are not measurable in claims data.

[b]We are exploring the feasibility of conducting an analysis of the length of time to nursing home placement.

[c]UNM has substantially increased its enrollment, but we have concerns that its patient eligibility criteria are substantially broader than anticipated. We are exploring the "effective sample size" and associated statistical power if we restrict analyses to beneficiaries with a more narrowly defined principal diagnoses of neuro-emergent conditions.

[d]We are currently negotiating with a large managed care organization to receive data to allow us to directly assess the feasibility of conducting a rigorous impact evaluation.

[e]Yale has substantially increased its enrollment, but we have concerns that it adopted a patient identification method and a definition of eligibility (based on risk of falling) that are not replicable in claims data. We are exploring the possibility of identifying beneficiaries at risk of falling through claims-based proxies.

There are a number of reasons why we cannot conduct a rigorous impact analysis for some awardees, as we describe in Table IV.2. Although there could be multiple reasons for some of these 15 awardees, we highlight for each awardee the reason that presents the most significant challenge and the one that is most unlikely to change. The discussion following the table summarizes the awardee-provided data (in addition to the core outcomes from claims data) that will be analyzed and reported. These awardee-provided data are of course restricted to participants and so cannot support an impact evaluation. Nonetheless, analysis of the data may provide context and insight into the awardees' programs and can provide evidence of whether outcomes are moving in a direction consistent with the program having impacts. We also identify the stakeholders (clinicians, staff, or beneficiaries) whose program experiences we will be reporting on, based on the surveys we conducted in 2017. The discussion in the awardee-specific narratives provides more in-depth analysis of the reasons that impact analyses cannot be carried out for each of these 15 awardees.

The wide variation across awardees in the minimum required number of treatment group cases for adequate precision (shown in Table IV.2, ranging from 176 for Four Seasons to 1,842 for University of North Carolina) arises for several reasons. The variance of expenditures differs somewhat across the awardees, though this is not the major reason for the different sample sizes required. More important factors include (1) the increase in treatment group sample sizes required when the analysis must use all those eligible for the program but only a fraction of eligibles participate (a participation rate of 50 percent requires a sample size four times larger than a program in which all eligibles participate) and (2) analyses using different cross-sectional samples for the baseline and follow-up periods require sample sizes that are roughly four times larger than those using the same individuals in the baseline and follow-up periods.

Table IV.2. Evaluation criteria for HCIA R2 awardees for whom a rigorous impact evaluation will not be conducted

	Projected Medicare enrollment with 6 months of program exposure	Projected Medicaid enrollment with 6 months of program exposure	Sample size of treatment group required for MDE of 20% for expenditures	Do claims identify the outcomes most likely to be affected?	Likelihood of solid comparison group	Primary reason for no rigorous evaluation
Youth with complex conditions						
BMC	n.a.	252	873	Some effects likely for observed outcomes, but important outcomes missing	Too early to determine due to delay in Medicaid data	Too few treatment beneficiaries
High-risk chronic conditions						
UCSD	1,176	214	704	Some effects likely for observed outcomes, but important outcomes missing	Serious concern	Lack of strong comparison group
Northwell	177	88	487	Yes	Serious concern	Too few treatment beneficiaries
FSCL	4,779	n.a.	176	Yes	Serious concern	Lack of strong comparison group
DMC	432	2,365	641	Yes	Serious concern	Lack of strong comparison group
Behavioral health and cognitive disorders						
Hopkins	249	n.a.	601	Some effects likely for observed outcomes, but important outcomes missing	Some issues but probably surmountable	Too few treatment beneficiaries
Amerigroup	n.a.	526	733	Some effects observed in claims data, but important effects likely missing	Too early to determine because of delay in Medicaid data	Most important outcome not included in claims data

Table IV.2 *(continued)*

	Projected Medicare enrollment with 6 months of program exposure	Projected Medicaid enrollment with 6 months of program exposure	Sample size of treatment group required for MDE of 20% for expenditures	Do claims identify the outcomes most likely to be affected?	Likelihood of solid comparison group	Primary reason for no rigorous evaluation
Acute and sub-acute conditions						
NHCHC	107	656	251	Yes	Serious concern	Lack of strong comparison group
UNM	780	408	953	Yes	Some issues but probably surmountable	Too few treatment beneficiaries
Icahn acute	184	218	305	Yes	Serious concern	Too few treatment beneficiaries
Icahn rehab	230	n.a.	242	Yes	Serious concern	Too few treatment beneficiaries
UMich	796[a]	0	291	Yes	Serious concern	Too few treatment beneficiaries
U NC	616	n.a.	1,842	Yes	Some issues but probably surmountable	Too few treatment beneficiaries
Primary and preventive care						
CHS	0	Not available	1,146	Yes	Too early to determine because of delay in Medicaid data	Lack of timely Medicaid data by final report
Wash U	0	1,526	1,146	Some effects likely for observed outcomes, but important outcomes missing	Too early to determine because of delay in Medicaid data	Lack of strong comparison group
Yale	1,186	281	359	Yes	Some issues but probably surmountable	Lack of strong comparison group

Note: These sample size requirements are for Medicare only or Medicaid only because the two groups of enrollees must be analyzed separately. Thus, an awardee for which neither sample alone exceeds the minimum but whose combined Medicare and Medicaid enrollees exceed the minimum sample size for the treatment group (for example, Icahn acute and UNM) still does not have enough enrollees to meet the precision standard.

[a]The enrollment total for UMich includes many enrollees whom we cannot include in the evaluation because they received a similar intervention in the pre-program period. See the discussion in the text below.

n.a. = not applicable.

Youth and young adults with chronic or complex physical conditions. Among this group of five awardees, we expect to be able to conduct a rigorous impact evaluation of four awardees' programs. The exception is Boston Medical Center, for which we do not expect to have sufficient statistical power because of low enrollment (252 Medicaid enrollees). We project that final enrollment will be only one-third of the number of Medicaid beneficiaries needed to conduct a rigorous impact evaluation. We will therefore report the experience of staff and participants from Mathematica-conducted surveys. Boston Medical Center is also conducting its own survey of program participants that will be analyzed by a local evaluator; and we are exploring the possibility of analyzing the data the awardee is collecting on its interactions with patients (such as in-person visits, electronic communications, and so on).

Adults with chronic conditions—high risk. Among this group of eight awardees, we expect to be able to conduct a rigorous impact evaluation of four programs. For Northwell, we do not expect to have sufficient statistical power because of low enrollment (177 Medicaid enrollees). For Four Seasons, the University of California at San Diego, and Detroit Medical Center, we have serious concerns about being able to draw a solid comparison group. Absent an impact analysis for Northwell, we will compare outcomes (hospitalization rates, use of home dialysis, and preemptive kidney transplantation) for program participants with national benchmarks drawn from the U.S. Renal Data System. We will also report the experiences of staff and participants from Mathematica-conducted surveys.

Enrollment into the Four Seasons program is based on a clinical assessment that is guided by a screening tool. The tool identifies significant risk factors such as physical limitations, serious illnesses, social determinants, and whether the provider would be surprised if the patient died in the next year. It is not possible to replicate this clinical assessment with claims and enrollment data. Therefore, we cannot ensure that a comparison group selected by using claims and enrollment data will be a proper counterfactual for the treatment group. We plan instead to draw a matched comparison sample for a subset of Four Seasons enrollees who died within one year of enrollment and had comparable diagnoses and characteristics. Difference-in-differences estimates derived from these samples will not yield estimates of program impacts for the full sample of enrollees but will indicate whether outcomes for the two groups of decedents differed systematically in the period before death. We will also report the experiences of staff and clinicians from Mathematica-conducted surveys.

For the University of California at San Diego, treatment beneficiaries are identified by using clinical data that are not available in claims data. The awardee is providing us with clinical information that it uses to generate baseline and follow-up cardiovascular risk scores for stroke and acute myocardial infarction. We evaluated the completeness of these data for use in an analysis of whether the awardee's program resulted in a change in cardiovascular risk over time. Because the follow-up clinical data are not complete, we cannot calculate change in risk score over the intervention period. We will report the experiences of staff and participants from Mathematica-conducted surveys.

For Detroit Medical Center, the concern about developing a solid comparison group is tied to the very low rate of participation among beneficiaries who were flagged as meeting the initial eligibility criteria—frequent ED use and no usual source of primary care—and who consented to

participate. To date, the awardee has identified almost 37,500 potential program participants based on a review of its EMR for frequent use of its EDs. However, less than 15 percent of potential participants have enrolled in the program[8] for three reasons: (1) they were not recruited before they left the ED and could not be contacted after discharge, (2) they were found to be ineligible because they stated that they had a primary care provider not identified in the awardee's EMR, or (3) they refused to enroll in the program. We have serious reservations about our ability to develop a claims-based algorithm that identifies such a small percentage of potential participants within an intent-to-treat framework, which will significantly dilute the intervention effects and will require a substantially larger sample than projected to detect a very small effect. We will attempt to devise a claims data algorithm—using observations on eligible and ineligible patients at the Detroit Medical Center—to assess the likelihood that participants were contacted, have a primary care provider, and agreed to participate. Given the small number of Medicare FFS beneficiaries, we will consider this approach for Medicaid-only beneficiaries and will document our results in the final report if Medicaid data become available and we are confident that we can produce unbiased program estimates. At this point, we do not have any awardee-specific data on implementation to report; however, we will report the experiences of staff and participants from Mathematica-conducted surveys.

Adults with chronic physical conditions—lower risk. We expect to be able to conduct a rigorous impact evaluation of all five awardees in this group.

Individuals with behavioral health and cognitive disorders. Among this group of five awardees, we expect to be able to conduct a rigorous impact evaluation for three of them. For Johns Hopkins University, which has only 249 Medicare enrollees, we do not expect to have sufficient statistical power to assess the effect of its program on expenditures. However, since the program is expected to affect expenditures by delaying the time to nursing home placement, we are exploring whether we can assess program effects on this outcome, since it may have smaller variance than expenditures, possibly yielding a detectable difference that is below our 20 percent ceiling. Absent an impact analysis, we will explore with the awardee the availability of data that it is collecting on participants who still reside at home at three points during their model testing period—baseline, 9 months, and 18 months. If the data are available and complete, we will report on changes over time in cognitive function, functional dependency, depression, patient's quality of life (at 18 months only), neuropsychiatric behavior, caregiver burden, and patient satisfaction with care (18 months only). We will also report the experiences of staff and participants from Mathematica-conducted surveys.

For Amerigroup, we do not expect to have sufficient statistical power because of low enrollment, but we are even more concerned that the primary outcomes of the program (such as health knowledge, life skills, and employment) are not observable in Medicaid claims data, and there is no other source of data available for any potential comparison group. The projected final sample size (526 enrollees) is not likely to be large enough to detect effects of 20 percent on

[8] In an earlier analysis of 20,000 potential program participants, patient navigators were able to contact only 40 percent at the time of the ED visit or shortly thereafter. Of the roughly 8,000 beneficiaries contacted, 60 percent were disqualified because they reported that they had a usual source of care that was not in the EMR. Of the remaining 3,000 eligible patients, 56 percent refused to participate.

most claims-based outcomes. However, we might have a large enough analysis sample to be able to detect whether Amerigroup affected the likelihood that participants would receive primary care and possibly whether they had any visits to the ED. In addition, the awardee is conducting surveys of participants that will include questions related to many of the outcome measures that are not available in the claims data. We will explore whether an analysis of the awardee-collected survey data would be worthwhile. We will also report the experiences of the awardee's staff from a Mathematica-conducted survey.

Adults who are seeking or who recently received acute or sub-acute care. Across the six groups of awardees, a rigorous impact evaluation is most challenging for awardees focusing on acute and sub-acute conditions. Of the six awardees in this group, we are likely to conduct a rigorous impact evaluation for one awardee, Mesa Fire and Medical Department. For the University of New Mexico, Mount Sinai (both the acute and rehabilitation components of its program), the University of Michigan, and the University of North Carolina, projected enrollment will not support detecting a 20 percent effect in total Medicare or Medicaid expenditures.[9] While we project a sizable number of enrollees in the University of Michigan program, and it exceeds the number needed to detect a 20 percent effect, the enrollment estimate overstates the number of beneficiaries that can be used in our analysis. Some of the beneficiaries in the finder file that we used to make the enrollment projections had actually participated in the program that predated HCIA R2 and therefore cannot be included in our evaluation. Although we do not plan to conduct an impact evaluation of Mount Sinai's intervention because of its small sample and the difficulty of selecting a proper comparison group, we do plan to contrast outcomes for the awardee's acute care treatment group with the outcomes for its selected comparison group. Although enrollment increased in the University of New Mexico's program, it still falls short of the number needed to detect 20 percent effects, and the variation in outcomes for enrollees is substantially greater than used in these minimum detectable effect calculations. We are examining the distribution of principal diagnoses for more recently enrolled participants to determine if it differs from that of participants who enrolled earlier, given the awardee's change in eligibility criteria and recruitment practices. For the University of Michigan, Mount Sinai, and the University of New Mexico, we will report the experiences of awardee staff, clinicians, and participants. For the University of North Carolina, we will report the experiences of awardee clinicians from our clinician survey.

For National Health Care for the Homeless, we are concerned about using the International Classification of Diseases (ICD) supplemental diagnosis codes that reflect homelessness to identify a solid comparison group that comprises homeless individuals with Medicare or Medicaid insurance.[10] Supplemental classifications are intended to identify factors other than

[9] The University of North Carolina enrolled over 600 Medicare beneficiaries in its program. Because our initial estimate of the variance of outcome measures for this program was quite high, this sample remains insufficient to detect a 20 percent effect. If our revised estimate of the variance is substantially lower than the initial value, we will reconsider the possibility of an impact evaluation.

[10] We will attempt to identify beneficiaries with a diagnosis of homelessness as follows: lack of housing (V60.0), inadequate housing (V60.1), economic problems (V60.2), other housing or economic circumstances (V60.89), or unspecified housing or economic circumstances (V60.9). The V codes are used to describe beneficiaries' encounters

disease or injury that constitute or contribute to health problems. Using 2014 Medicare claims data, we identified 7,869 beneficiaries with a homeless diagnosis code in the five study states, which confirms that some providers do bill by using these supplemental diagnosis codes. We will evaluate the rate of reporting of homelessness on Medicare and Medicaid claims for the treatment group and will revisit this assessment of evaluability should there be a moderate to high rate of claims-based identification of the treatment group as homeless. An initial evaluation of Medicaid data for two of the five states shows that the reporting of homelessness in Medicaid encounter data is poor, so we are not optimistic. We will continue to report on the awardee's self-monitoring measures to examine the implementation of the core elements of respite care—that is, care management, patient engagement, and transitional care. We will also report the experiences of staff from a Mathematica-conducted survey.

Individuals needing primary or preventive care. Among this group of eight awardees, we expect to be able to conduct a rigorous impact evaluation for five of them. For Yale University, we are restricted to analyzing one of the three program arms: treatment group beneficiaries who had an ED visit for a fall; thus, we do not believe that projected enrollment in this treatment arm will support detecting a 20 percent effect in total Medicare or Medicaid expenditures despite an increase in recent enrollment. It is our understanding that recruitment now includes beneficiaries who are at risk of falling, but at this point in time, we are not certain that we can identify this population with claims data. We are currently looking into whether we can develop a claims-based proxy for risk of fall that captures a high proportion of program enrollees. We will monitor enrollment and reassess evaluability should enrollment exceed our current expectations. Absent an impact evaluation, we will report on the rate of receipt of intervention services (e.g., a paramedic visit and a Visiting Nurse Association visit) from participant-level data that the awardee has provided and from the data on the experiences of awardee staff and participants from Mathematica-conducted surveys.

For the Children's Home Society, it is unlikely that we will be able to conduct a rigorous impact evaluation because the awardee did not provide patient identifiers that allow us to link their records to Medicaid data. Instead, the awardee provided identifiers that allow us to link to each Florida Medicaid managed care organization that is under contract with the state to provide Medicaid managed care services. We are in discussions with the managed care organization that has the largest share of the awardee's participants. If we are able to negotiate claims data for the treatment group, and for a pool of the organization's insured who are not receiving the intervention and can be used for our comparison group, we will reassess evaluability. If managed care encounter data are not available, we will not conduct an impact analysis and will report only the experiences of awardee staff and participants from Mathematica-conducted surveys.

For Washington University, there are significant challenges that will limit our ability to construct a strong comparison group. In particular, for any pre-period analysis, we cannot identify the members of the treatment and potential comparison groups who were or were not in the Contraceptive CHOICE project, which provided services to a large number of women until 2013 and on which the current contraceptive program was modeled. We are concerned that the

with circumstances other than disease or injury, including circumstances that can influence a person's health status, such as disease exposures or, in our case, homelessness.

CHOICE program has likely contaminated the pool of participants and the baseline outcomes in Missouri. In addition, there are no comparable data sources on key outcomes for the comparison group or for the pre-enrollment period for the treatment group. We are also concerned that many women who have Medicaid insurance during the program period may not have had it in the pre-period, given that the program focuses partly on moving uninsured women onto Medicaid, which makes it difficult to track individuals over time in claims data and to isolate the impact of the Contraceptive Choice Center alone. Consequently, we propose to carry out aggregate comparisons of key outcomes over time between program participants, women in Missouri and Illinois more broadly, and women in other geographic regions. To conduct this analysis, we will use nationally available survey data, including data from the National Survey of Family Growth (NSFG), the Behavioral Risk Factor Surveillance System (BRFSS), and the Pregnancy Risk Assessment Monitoring System (PRAMS). Further, using awardee-provided data, we will report on the use of long-acting reversible contraception, on unintended pregnancies, and on births from unintended pregnancies among the treatment group beneficiaries and between subgroups of treatment group beneficiaries. We will also report the experiences of awardee staff, clinicians, and participants from Mathematica-conducted surveys.

Overall, we are reasonably confident that we can establish a credible comparison group to support a difference-in-differences analysis for the majority of the awardees' programs. We will continue to refine our strategies for estimating program impacts by addressing the distinct challenges presented by each awardee with respect to identifying a credible comparison group. As data become available, it will be necessary to ascertain how well we can match beneficiaries and providers on observable characteristics.

D. Characteristics of Medicare and Medicaid participants at baseline

The HCIA R2 awardees have implemented programs directed at widely varying groups of Medicare and Medicaid beneficiaries. Some programs addressed complex medical conditions in young people. Others targeted chronic conditions, behavioral health, or cognitive disorders. Still others were aimed at aspects of acute care, long-term care, or primary and preventive care. By mid-2017, more than 150,000 Medicare beneficiaries and at least 15,000 Medicaid beneficiaries had participated, either directly or indirectly, in the 38 programs under study.

All 38 awardees have submitted to us at least one file that identifies program participants who are enrolled in Medicare or Medicaid; these files are known as "finder files". Nine awardees are not considered in this report for one of two reasons.

1. Medicaid data covering the program period are not yet available for six awardees (Amerigroup, Boston Medical Center, Clifford Beers Guidance Clinic, Seattle Children's Hospital, Washington University, and Columbia University).
2. Three awardees will provide Medicaid data to us directly but have not done so in time for this report (Altarum, the Children's Home Society, and the National Association of Children's Hospitals).

Table IV.3 shows awardees included and excluded from the current analysis of baseline characteristics, the reasons for exclusion, and the number of participants who met the inclusion

criteria. The table further identifies the five awardees for which we present a comparison or control group and the three for which we present preliminary impact estimates. Baseline totals and characteristics are presented as of May 31, 2016 for most awardees but have been updated to later dates for 11 awardees for which a comparison group is presented or for those that had submitted their final finder files by mid-2017.

There is marked variation across awardees in the number of Medicare FFS beneficiaries enrolled in the awardee's program through May 2016 or later (cutoff dates ranging from November 2016 through August 2017 were used for 11 awardees), and who met the inclusion criteria. Seven had over 1,000 such beneficiaries. Another seven had fewer than 300, an amount that we and many other evaluators view as the minimum number required to support a formal impact analysis of claims-based outcomes.[11] However, these counts may increase substantially for some awardees by the end of their intake period.

Despite wide variation in health status and spending among program participants, taken as a group, they appear to be less well-off and have a greater need for care than most Medicare beneficiaries. Table IV.4 presents summary characteristics for Medicare beneficiaries enrolled in programs operated by the 27 awardees for which we could conduct an initial baseline analysis. Overall, beneficiaries were 10 percentage points more likely than the national Medicare population to have qualified for Medicare on the basis of a disability and nearly twice as likely to be enrolled in Medicare and Medicaid. Their mean Medicare spending in the year prior to enrollment was more than twice the national average.

Table IV.3. Awardees included in and excluded from the baseline analysis

Awardee	Received usable finder file in time for inclusion?	Cumulative number of eligible Medicare FFS beneficiaries	Awardee included in this report	Comparison group included in this report	Impact estimates included in this report
All awardees			29	5	3
Youth with complex medical conditions					
BMC	Yes, but mainly Medicaid enrollment	n.a.	N	N	N
NACHRI	Yes, but mainly Medicaid enrollment	n.a.	N	N	N
SCH	Yes, but mainly Medicaid enrollment	n.a.	N	N	N
UIC	Yes, but mainly Medicaid enrollment	n.a.	Y[a]	Y[a]	Y[b]
WI DHS	Yes, but mainly Medicaid enrollment	n.a.	Y[a]	N	N
High-risk chronic conditions					
DMC	Yes	162	Y	N	N
FSCL	Yes	2,050	Y	N	N
NM	Yes	351	Y	Y	N
Northwell	Yes	97	Y	N	N
NYC H+H	Yes	3,015	Y	Y	Y
UCSD	Yes	1,130	Y	N	N
UHCMC	Yes	327	Y	N	N
U KS	Yes	1,244	Y	N	N

[11] The threshold of 300 assumes a minimum detectable effect of 20 percent of the mean for an outcome variable with a coefficient of variation (the variable's standard deviation divided by its mean) of 1.0, such as a binary variable with mean of .50.

Table IV.3 *(continued)*

Awardee	Received usable finder file in time for inclusion?	Cumulative number of eligible Medicare FFS beneficiaries	Awardee included in this report	Comparison group included in this report	Impact estimates included in this report
Lower-risk chronic conditions					
ACCF	Yes	918	Y	N	N
CCNC	Yes	40,480	Y	N	N
CHIIC	Yes	4,678 [c]	Y	N	N
FPHNY	Yes	350	Y[a]	Y	N
Ventura	Yes	416	Y	N	N
VillageCare	Yes	418	Y	N	N
Behavioral health or cognitive disorders					
Amerigroup	Yes, but mainly Medicaid enrollment	n.a.	N	N	N
Clifford Beers	Yes, but mainly Medicaid enrollment	n.a.	N	N	N
Hopkins	Yes	249	Y	N	N
Montefiore	Yes	122	Y[a]	N	N
UCSF	Yes	557	Y	Y	Y[b]
Acute and sub-acute care					
Icahn (HaH & RaH)	Yes	414	Y	N	N
Mesa	Yes	629	Y	N	N
NHCHC	Yes	61	Y	N	N
UMich	Yes	222	Y	N	N
UNC	Yes	228	Y	N	N
UNM	Yes	184	Y	N	N
Primary or preventive care					
AAMC	Yes	91,064	Y	N	N
Altarum	Yes, but mainly Medicaid enrollment	n.a.	N	N	N
Avera	Yes	5,444	Y	N	N
CCC	Yes	1,425	Y	Y	N
CHS	No	n.a.	N	N	N
Columbia	Yes, but mainly Medicaid enrollment	n.a.	N	N	N
Wash U	Yes, but mainly Medicaid enrollment	n.a.	N	N	N
Yale	Yes	193	Y	N	N

Note: Reporting dates for the cumulative number of Medicare FFS beneficiaries vary across awardees as follows: Avera, Hopkins, NM, NYC H+H, UCSF, VillageCare (November 30, 2016); FPHNY, UCSD (February 28, 2017); CCC (April 30, 2017); CHIIC (May 31, 2017); Icahn (August 31, 2017). All others (May 31, 2016).

[a]We present baseline demographic, expenditure, and service use characteristics for Medicare beneficiaries for 27 awardees, and for 3 of these 27 awardees (FPHNY, Montefiore, and VillageCare), we also present demographic, expenditure, and service use characteristics for Medicaid beneficiaries. For two awardees (UIC and WI DHS), we do not present expenditure or service use characteristics because they do not have any Medicare beneficiaries, and their Medicaid claims data were not available for the baseline period. For WI DHS, we present only demographic characteristics and only for Medicaid beneficiaries.

[b]For UIC and UCSF, we are using participants randomly assigned to treatment and control groups to measure program impacts.

[c]CHIIC provided encounter-level data, with one record for each encounter with a health coach. The encounter data include the participant's name, date of birth, and gender, as well as date and type of health coaching service. However, the awardee did not collect or provide beneficiary IDs or payer information. As a result, we identified participants in Medicare enrollment data using their name, date of birth, gender, and state of residence. For those participants for whom we could find a match, we used their Medicare IDs to pull their claims.

FFS = fee-for-service; n.a = not applicable; N = no; Y = yes.

Table IV.4. Selected characteristics of Medicare FFS beneficiaries when they enrolled in the awardees' programs

Awardee	Cumulative number of eligible Medicare FFS[a]	Percentage of participants younger than 65	Percentage of participants who originally qualified for Medicare on the basis of disability	Percentage of participants dually eligible for Medicare and Medicaid	Mean HCC score[b]	Average Medicare expenditures PBPM
Medicare population[c]	—	17%	24%	18%	1.00	$792
All awardees[d]	156,081	25%	34%	34%	2.04	$2,109
High-risk chronic conditions						
DMC	162	61%	71%	51%	1.50	$1,503
FSCL	2,050	<1%	15%	25%	3.46	$3,026
NM	351	32%	45%	30%	1.96	$1,745
Northwell	97	16%	29%	15%	2.66	$2,353
NYC H+H	3,015	31%	41%	54%	1.52	$2,045
UCSD	1,130	9%	17%	19%	1.15	$867
UHCMC	327	14%	20%	14%	3.74	$3,719
U KS	1,244	11%	21%	16%	1.74	$1,382
Lower-risk chronic conditions						
ACCF	918	10%	20%	12%	1.48	$986
CCNC	40,480	41%	55%	72%	1.47	$978
CHIIC	4,678	10%	18%	15%	1.36	$1,071
FPHNY	350	41%	51%	56%	2.22	$3,021
Ventura	416	51%	66%	80%	1.58	$1,262
VillageCare	418	86%	90%	65%	1.66	$1,593
Behavioral health or cognitive disorders						
Hopkins	249	6%	22%	70%	1.99	$1,892
Montefiore	122	43%	51%	38%	1.47	$1,754
UCSF	557	2%	7%	13%	1.34	$1,041
Acute and sub-acute care						
Icahn (HaH)	184	7%	15%	30%	2.75	$3,167
Icahn (RaH)	230	3%	12%	17%	3.52	$5,571
Mesa	629	29%	38%	30%	2.24	$2,549
NHCHC	61	77%	79%	79%	2.70	$4,357
U NC	228	18%	28%	20%	0.95	$576
UMich	222	22%	32%	18%	2.32	$2,188
UNM	184	18%	28%	30%	1.86	$1,530
Primary or preventive care						
AAMC	91,064	19%	25%	17%	1.28	$1,109
Avera	5,444	8%	19%	23%	2.46	$2,245
CCC	1,425	10%	16%	12%	2.48	$2,509
Yale (ED visit)	193	10%	22%	34%	2.32	$3,004

Source: Medicare claims and enrollment data, September 1, 2013–varying dates. For AAMC, ACCF, CCNC, DMC, FSCL, Mesa, Montefiore, NHCHC, Northwell, U KS, U NC, UHCMC, UMich, UNM, Ventura, and Yale, data are through May 31, 2016. For Hopkins, data are through September 30, 2016, and for Avera, NM, NYCH+H, UCSF, VillageCare data go through November 30, 2016. For FPHNY and UCSD, data are as of February 28, 2017. For CCC, data are as of April 30, 2017 and for CHIIC, data are as of May 31, 2017. Finally, data for Icahn are as of August 31, 2017.

[a] Beneficiaries included in this table were enrolled in an awardee's program on or before May 31, 2016. Beneficiaries must have been enrolled in Medicare FFS with both Part A and B coverage, and have Medicare as primary payer on their enrollment date and for at least 90 days during the 365 days before their enrollment date.

[b] HCC scores were calculated by Mathematica for the same 365-day period.

[c] National averages are based on the 5% Medicare FFS sample for 2014.

Table IV.4 *(continued)*

[d] Mean values in this row are the means across awardees (i.e. the means of the numbers in the column), not the means over the 156,081 beneficiaries in the HCIA R2 program.

ED = emergency department, FFS = fee-for-service, HaH = Hospital at Home; HCC = hierarchical condition categories; PBPM = per beneficiary per month; RaH = Rehabilitation at Home.

We have processed Medicaid data for five awardees. Table IV.5 compares participants in the five programs across four dimensions depending on whether appropriate data were available to us. With the exception of the University of Illinois, each of the awardees enrolled a few hundred participants, nearly all of whom are eligible for full Medicaid benefits. For the three awardees for which information was available, the proportion of participants who were also enrolled in Medicare is strikingly similar across programs, at about 24 percent. The much lower mean Medicaid expenditures for Montefiore Medical Center than for the Fund for Public Health in New York and VillageCare is due to differences in their target populations. Montefiore Medical Center targeted people with behavioral health problems, including children with attention deficit hyperactivity disorder. VillageCare targeted people with HIV, and the Fund for Public Health in New York targeted people with hepatitis C; these latter two patient groups have very high prescription drug costs.

Table IV.5. Characteristics of program participants enrolled in Medicaid for five awardees

Awardee	Cumulative number of eligible Medicaid FFS enrollees as of May 31, 2016	Percentage with full Medicaid benefits	Percent dually eligible for Medicare and Medicaid	Mean Medicaid spending in baseline year
FPHNY	552	97%	25%	$4,096
Montefiore	487	96%	21%	$1,194
UIC	3,131[a]	NA	NA	NA
VillageCare	530	98%	24%	$5,445
WI DHS	203	NA	NA	NA

[a] UIC randomized 3,131 participants into treatment and control groups. The total UIC program enrollment was more than 14,000.

FFS = fee-for-service; NA = not available.

E. Synthesis of impact evaluations

1. Methods

The three interim impact evaluations presented in this report include two randomized control trials (RCTs) and one quasi-experimental analysis. For the two RCTs, beneficiaries were randomly assigned to either a treatment group or a control group. For the quasi-experimental analysis, we selected a comparison group to match the awardee's treatment group beneficiaries based on a pre-specified set of eligibility criteria. For this awardee, we applied optimal propensity score matching methods (Rosenbaum 1991) to create a comparison group of beneficiaries who were most similar to treatment beneficiaries on observable characteristics available from claims data. The overall quality of the matched comparison group constructed for New York City Health + Hospitals was excellent. Mean values of variables strongly related to outcomes (for example, prior hospitalization and Medicare spending) were quite similar for

treatment and comparison groups.[12] We used the randomized control groups or the propensity score matched comparison groups to model the counterfactual—our estimate of the mean outcomes that would have been observed for the treatment group beneficiaries absent the HCIA R2 interventions.

Our framework is an intent-to-treat approach—meaning that all enrolled beneficiaries are retained in the sample regardless of whether they actually participate or whether they drop out of the program—and we follow the treatment and control/comparison beneficiaries over time. We used Medicare FFS claims, Medicaid encounter data, and Medicare and Medicaid enrollment data from 2013 to 2017 to identify the treatment and comparison beneficiaries and to construct beneficiary sociodemographic and service use characteristics. We constructed the patient characteristics variables for a 12-month baseline period and the outcome variables for a 12-month follow-up period for our primary analysis as well as shorter 6-month periods and up to 24 months for those awardees with sufficient sample size. The length of follow-up varies across awardees and is primarily driven by the number of beneficiaries for which we have claims data covering that interval. We conducted the impact analyses for the University of Illinois on Medicaid beneficiaries and the other two impact analyses on Medicare beneficiaries.

We used regression models to estimate program effects. For the two RCT interventions, we used regression models to estimate mean differences in outcomes between the treatment and control groups during the intervention period, controlling for baseline beneficiary characteristics and baseline values of the outcomes. For the quasi-experimental analysis, we used a difference-in-differences regression model to estimate the covariate-adjusted pre-post changes in outcomes for beneficiaries in the treatment group and for the matched comparison group over the same period. These models for the quasi-experimental designs control for all baseline beneficiary characteristics used in selecting the matched comparison group, with the exception of the baseline outcome variables. The effects of those variables are captured through the difference-in-differences estimation model. For these quasi-experimental designs, we estimated the program impact as the difference between the average change over time for treatment beneficiaries and the average change over the same period for the matched comparison beneficiaries.

We examined the impact of the interventions on several key outcomes of interest that can be broadly grouped into three categories: (1) Medicare (Medicaid) expenditures in total, and by major types of services; (2) rates of service use; and (3) likelihood of acute care utilization. We analyzed a common core set of outcome measures across awardees when applicable to the intervention. We also evaluated effects on some awardee-specific measures that are only applicable to a given intervention. Note that sample sizes in the impact analysis below differ from those in other tables, due to the need to limit the sample to beneficiaries with at least six months of follow-up data.

[12] We have also selected comparison groups for two other awardees, University of Nebraska and CareChoice, and found those matches to be very strong as well. However, those comparison groups were not selected in time for us to include preliminary impact analyses in this report. The quality of the matches for all three of these comparison groups is explained in detail in Appendix B.

2. Findings

Table IV.6 displays the estimates that are significantly different from zero and an indicator (N/S) of the other tests that were conducted for which the estimated effect was not significantly different from zero at the 0.10 level.[13] The one exception is for the outcome of total per beneficiary per month (PBPM) Medicare or Medicaid expenditures. We included these impact estimates for all three awardees even though none are statistically significant to show the magnitude and direction of this overall summary measure of program effects on utilization and expenditures. The impact estimate also provides a point of comparison to the calculations we present indicating how large the estimated impact would have had to be for us to have confidence that the program had a real effect (that is, the size required for statistical significance, given the estimated standard error). The estimated effects of the three awardees' programs on mean expenditures and rates of utilization are all expressed as a percentage of our estimate of what the mean outcomes would have been for participants absent the intervention. The estimated effects on the likelihood of any acute care utilization are expressed as percentage point changes. We display results for each awardee's full study sample for a 12-month follow-up period. For some awardees—those with large enough samples—we also estimated impacts for key subgroups, and when relevant to the intervention and possible with available data, we also examined impacts with shorter and/or longer follow-up periods. We do not report these results in Table IV.6, but we do reference relevant findings in the text in this section and provide more detailed reporting of the results in each individual awardee report.

For the three awardees, these early findings from our impact evaluation are mixed, and they differ substantially across the awardees and across outcome measures. None of the programs has a statistically significant effect on total expenditures. Of the 33 outcomes for which we estimated program effects on the full evaluation samples for these awardees, 7 (about one-fifth) are significantly different from zero at the 0.10 level or better. For one of the three awardees, our estimates provide evidence that its intervention is having favorable effects on ED visits. However, for one awardee we found one outcome for which we estimated an unfavorable intervention effect, an increase in ED visits. For the third awardee, our estimates suggest that several outcomes are better than they would have been absent the intervention, but none of the estimates are statistically significant, perhaps because this awardee's sample is small. Below we summarize the key findings for these three awardees.

The University of California at San Francisco is testing a model of dementia care that addresses the unmet needs of community-dwelling patients and caregivers by providing additional telephone-based supportive care and education, as well as medication consultation and support in planning future medical, financial, and legal decisions. The awardee hypothesized that giving patients and caregivers personalized preventive care over the phone should reduce the program participants' incidence of medical emergencies, prevent unnecessary ED and hospital use, delay nursing home placement, and result in overall cost savings. We evaluated a small

[13] The designation N/S, which indicates that the estimated effect is not statistically significant at the 0.10 level, does not imply that the program had no effect. Rather, it indicates that given the precision that we have for these estimates, the difference between the two groups is within the range that might reasonably occur simply by chance rather than because of a true program effect. The statistically significant estimates lie outside the range that one would expect to observe by chance.

subset of outcomes given the small number of beneficiaries included in this interim analysis, 247 treatment and 132 control beneficiaries, and found no robust evidence of program effects. However, the sample is not large enough for us to be confident of detecting even a 20 percent effect of the program on total Medicare FFS expenditures, rates of service utilization, or the likelihood of any acute care utilization. Given the standard error of the impact estimate, the estimated effect on total expenditures would have had to be 26 percent or more of the treatment group's predicted mean in order to be statistically significant. Our final analysis will include all 539 participants in the program, which will improve our precision and should reduce the detectable effects by about one-third. However, it will still be the case that unless estimated program impacts are large, then the impact estimates will not be statistically significant.

Table IV.6. Summary of percent effect of three awardees' programs on expenditures, rates of utilization, and likelihood of any acute care utilization during a 12-month follow-up period

Expenditures and utilization measures	UCSF (247/132)	UIC[a] (3,131/3,128)	NYC H+H (2,042/6,145)
Comparison group	**RCT**	**RCT**	**Matched group**
Medicare or Medicaid expenditures			
Total	-8% N/S	-8% N/S	3% N/S
Acute inpatient	N/S	N/S	N/S
Inpatient other			N/S
Outpatient			N/S
Physician services			-10%***
Home health			N/S
SNF			N/S
Hospice			N/S
Durable medical equipment			N/S
Outpatient ED		N/S	
Prescription drugs		N/S	
All other Medicaid-covered services		N/S	
Service utilization rates			
Acute hospital admissions	N/S	N/S	N/S
Outpatient ED visits	N/S	-7%*	N/S
Primary care visits in ambulatory settings			-13%***
Specialist visits in any setting			-10%***
Specialist visits in ambulatory settings			-11%***
Percentage of beneficiaries with a service use			
Percentage with a hospitalization	N/S	N/S	N/S
Percentage with an outpatient ED visit or observation stay	N/S	-2.4*	+2.2*
Percentage with a readmission among all beneficiaries	N/S		N/S

Note: Percent effect = Impact regression coefficient/(Predicted treatment group mean – impact regression coefficient)*100. N/S = estimate not statistically significant; shaded cells = not evaluated.

Treatment and comparison group sample sizes (n_T/n_C) appear in parentheses under the awardee names.

[a] For UIC, the outcomes are Medicaid expenditures per beneficiary per month.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

The University of Illinois' CHECK program was intended to improve care coordination for children and young adults who have chronic medical conditions by addressing both health care needs and social needs. The awardee hypothesized that improved access to social services and to primary, specialty, and mental health care will result in better health and social outcomes, including fewer hospitalizations and ED visits, as well as lower costs. Our findings show that CHECK reduced outpatient ED visits by 7 percent and the likelihood of an ED visit by 2.4 percentage points. Most of that reduction in utilization occurred within the first six months of when beneficiaries entered the intervention. In addition, CHECK reduced Medicaid *inpatient* expenditures for medium- and high-risk subgroup of beneficiaries (not reported in Table IV.6). This reduction was driven by reductions in both "any" inpatient hospital care and the number of hospitalizations among the treatment group during the second half of the first RCT year. Although we observed 8 percent lower total PBPM Medicaid expenditures, the estimated effect would have had to be 15 percent or greater of the treatment group's predicted mean in order to be statistically significant. For the medium- and high-risk subgroups in months 7 to 12 (not shown, see University of Chicago narrative), we found evidence of reduced PBPM Medicaid hospital spending and admissions, and a reduction in the percentage of participants with any outpatient ED visits.

New York City Health + Hospitals' ED Care Management Initiative provides comprehensive care management to patients with ambulatory care sensitive conditions for 90 days following ED discharge by increasing linkages to primary care and specialty providers. The program's primary expected effect is a reduced incidence of medical emergencies that would otherwise lead to preventable ED and hospital use. Our findings suggest that the program did not reduce ED use as intended; instead, the percentage of treatment beneficiaries with an ED visit increased 2.2 percentage points above what would have been predicted absent the intervention. However, for beneficiaries with asthma, the average number of ED visits was 22 percent lower than the predicted value absent the intervention. Furthermore, we found unexpected evidence that the program led to 10 percent and 13 percent fewer physician visits for primary care and specialist care, respectively, with an accompanying 10 percent reduction in average physician expenditures. This decrease in ambulatory care use might be contrary to the intended goals of the program, which were to strengthen contact between beneficiaries and ambulatory health care services. However, the decrease may also suggest that other services that the nurse care manager provides, such as health education and referrals to social workers, may reduce the need for ambulatory care from physicians. These mixed results for ED and primary care visits roughly offset each other: we estimated a 3 percent increase in total PBPM Medicare expenditures, which is not statistically significant. The estimated effect would have had to be 8 percent or greater of the treatment group's predicted mean in order to be statistically significant.

3. Discussion

These conclusions should be regarded as highly preliminary because the treatment groups that we analyzed consist of only the beneficiaries who entered the awardees' programs during the first two years of the cooperative agreement. Thus, the estimates cover only the early period of program operations, when the awardees were still learning how best to implement their interventions and focusing substantial resources on identifying and enrolling patients, and on engaging both providers and patients. Beneficiaries also had relatively short exposure to the

programs over this time frame. Finally, the matching process will be updated for these awardees to include more recent enrollees and to reflect lessons learned in this initial matching effort, so the comparison groups will change. In subsequent analyses, we will re-estimate the models to arrive at final impact estimates for the life of each program, for a larger study sample, and for different intervals of follow-up time; we will also perform more extensive sensitivity and robustness checks on all results. In addition, we will estimate program impacts in a Bayesian framework, to increase precision and present probabilities that true effects exceed some meaningful threshold.

V. FINDINGS FROM EVALUATION OF THE PAYMENT MODELS

A. Introduction and summary of findings

As part of their HCIA R2 applications, the awardees' were required by CMS to submit an initial payment model design that would support their proposed service delivery models. CMS also expected awardees to refine and submit a final payment model design by the end of the three-year cooperative agreement (August 2017). In doing so, CMS's goals were to (1) have awardees design models that could provide a sustainable source of funding for service delivery after the cooperative agreement ended and (2) be able to identify promising models that might justify re-testing or spreading them in the future.

This chapter (1) categorizes the payment models by level and type of financial risk, quality of care requirements, and incentives for providers; (2) summarizes the awardees' progress in developing their payment models, (3) assesses factors associated with this progress, and (4) discusses the implications for meeting the HCIA R2 goals and for similar programs in the future.

We describe and assess the payment models based on our analysis of (1) the awardees' final payment model reports submitted in June 2017, (2) prior payment model and narrative reports from the awardees as needed, and (3) semi-structured telephone interviews with a select subset of awardee leaders that we conducted in August through October 2017. The interviews were focused specifically on the payment models, and we spoke with awardee leaders and/or the leaders at the implementing sites who met one or more of the following criteria: (1) had advanced to the negotiation stage with at least one payer, (2) had implemented a payment model during the cooperative agreement, or (3) had a model that was of special interest to CMS. The interview goals were to clarify and expand on the details of the models that had advanced past the development stage.

In summary, we identified 46 different payment models proposed by 37 awardees.

- Most models incorporated one or more aspects of an alternative payment model (APM), moving away from FFS and toward value-based purchasing.
- Eight awardees executed a contract with at least one payer, and an additional 15 were in negotiations with payers.
- Common factors associated with awardees who made progress on their payment models included (1) consulting with experts, (2) leveraging existing APM initiatives, (3) collaborating with payers, (4) learning from peer groups, and (5) acquiring data to calculate costs and demonstrate value.
- We did not detect a meaningful relationship between the awardees' progress in developing their payment models and overall implementation effectiveness.

B. Categorizing the payment models in an APM framework

Based on the awardees' reports and on our interviews with awardee leaders, we identified 46 payment models among 37 awardees.[14] Five awardees developed multiple payment models to meet the unique needs of each implementation site, support different levels of service delivery, and/or to engage multiple payers (National Association of Children's Hospitals, National Health Care for the Homeless, New York City Health + Hospitals, the University of California at San Francisco, and Cleveland Medical Center). For these five awardees, we analyzed each model separately.

To better understand the awardees' progress in developing alternative payment models (rather than just continuing the FFS model), we categorized all 46 of the models from 37 awardees according to the APM framework developed by CMS and refined by the Health Care Payment Learning & Action Network, most recently in 2017.[15] The APM framework groups payment models into four main categories along a value-based continuum from FFS to capitated, population-based approaches:

- **Category 1**. FFS with no links to quality or value
- **Category 2**.Traditional FFS payments plus one of the following:
  - Subcategory 2A. Foundational payments to fund services that have the potential to improve quality and value (such as PBPM care coordination fees) but that offer no direct incentives to providers
  - Subcategory 2B. Provider incentives for reporting measures of quality or value
  - Subcategory 2C. Provider incentives for performance on measures of quality or value
- **Category 3.** Traditional FFS payments plus some form of provider financial risk, including the following:
  - Subcategory 3A. Shared savings with quality targets
  - Subcategory 3B. Shared savings and downside risk with quality targets
  - Subcategory 3N. Risk-based payments not linked to quality
- **Category 4.** Population-based payments, including the following:
  - Subcategory 4A. Payments based on specific conditions or scopes of practice (such as a comprehensive cancer care bundle)
  - Subcategory 4B. Prospective, population-based payments covering all of an individual's health care needs

[14] Amerigroup is not included in the analysis because the service delivery model will be supported by a state grant after the cooperative agreement ends, and the awardee did not propose another payment model.

[15] Health Care Payment Learning & Action Network. Alternative Payment Model (APM) Framework: Refreshed for 2017. McLean, VA: The MITRE Corporation. 2017. Available at: https://hcp-lan.org/groups/apm-fpt-work-products/. Accessed on: December 13, 2017.

- Subcategory 4C. Same type of payments as Subcategory 4B but within an organization in which the payer and delivery system components are integrated
- Subcategory 4N. Capitated payments without any minimum expectations or links to quality

For the purposes of this analysis, we focused on payment models that involved an agreement between a payer (Medicaid, Medicare, commercial insurers, and/or managed care plans associated with each of these payers) and an organization that supports the delivery of program services. This organization could be either a direct service provider or an organization that contracts with direct service providers, such as an accountable care organization (ACO). We did not categorize awardees based on payment approaches that the service delivery organization might take with its employed or contracted providers. For example, we categorized the payment models proposed by the National Association of Children's Hospitals' sites on the basis of the agreements between state Medicaid agencies or Medicaid managed care organizations (MCOs) and the children's hospitals that are the direct providers of program services. We did not assess how the hospitals pay or incentivize practices that participated in the awardee's program.

The majority of the awardees' payment models incorporated an APM approach beyond traditional FFS (Table V.1). About one-third of the 46 payment models were FFS linked to quality and value (category 2), and about 1 in 5 were in each of the two categories that have stronger incentives and risks (APMs built on FFS and population based payments; 36 of the 46 models). Ten models were FFS with no incentives to enhance quality or value. Overall, about 30 percent (14) of the contracts involved a risk of financial loss; 5 of the models built on FFS included such downside risk, and the 9 population-based payments by definition included such risk either for just the treatment of specific conditions or for all care. The awardees for 8 of these 14 risk-bearing models reached a contract with payers, and one awardee, National Care for the Homeless, accounted for 3 of these contracts.

Table V.1. Payment models by APM framework category

Category 1: FFS not linked to quality & value	Category 2: FFS linked to quality and value	Category 3: APMs built on FFS architecture	Category 4: Population-based payment
1A. FFS AAMC (S1) CCC (S1) Clifford Beers (S3) DMC (S1) Mesa (S1) NHCHC site 2 (S3) NHCHC site 5 (S3) UCSF model 1 (S3) UMich (S2) UNM (S2)	**2A. Foundational payments for infrastructure and operations** CHS (S1) Columbia (S1) Hopkins (S1) NACHRI site 2 (S3) NACHRI site 3 (S3 UCSD (S2) UIC (S2) Wash U (S2) WI DHS (S2) VillageCare (S1) **2B. Pay-for-reporting** None **2C. Pay-for-performance** Altarum (S1) CCNC (S2) Montefiore (S2) NYCH+H model 2 (S2) SCH (S2) Ventura (S1)	**3A. APMs with shared savings with no downside risk** ACCF (S1) BMC (S2) CHIIC (S3) UCSF model 2 (S3) UHCMC model 1 (S1) **3B. APMs with shared savings and downside risk** FPHNY (S2) NM (S1) Icahn (S3) U NC (S2) **3N. Risk-based payments NOT linked to quality** NACHRI site 1 (S3)	**4A. Condition-specific population-based payment** Avera (S2) FSC (S1) NHCHC site 1 (S3) NHCHC site 3 (S3) NHCHC site 4 (S3) Northwell (S2) UHCMC model 2 (S3) **4B. Comprehensive population-based payment** NYCH+H model 1 (S3) U KS (S2) **4C. Integrated finance & delivery system** None **4N. Capitated payments NOT linked to quality** Yale (S1)

Notes: Amerigroup is not included in the table because the service delivery model will be supported by a state grant after the cooperative agreement ends, and the awardee did not propose another payment model

S1 = stage 1, defined as model in development, and awardee not in active negotiations with payer(s)

S2 = stage 2, defined as awardee in active negotiations with payer(s)

S3 = stage 3, defined as awardee has a contract with at least one payer

APM = alternative payment model; FFS = fee-for-service; health IT = health information technology.

C. Awardees' progress in developing payment models

We described the awardees' progress in developing payment models by classifying the 46 models into one of three stages of development: (1) in development, but the awardee is not in active negotiations with any payers, (2) developed, and the awardee is in active negotiations with one or more payers, and (3) developed, and the awardee has a contract with at least one payer (Table V.2).

Table V.2. Models by stage of development at the end of the third year of the cooperative agreement

Stage 1: Model in development and awardee not in active negotiations with payer(s) (N = 15 models, 15 awardees)	Stage 2: Awardee in active negotiations with payer(s) (N = 16 models, 16 awardees)	Stage 3: Awardee has a contractual agreement with at least one payer (N = 15 models, 8 awardees)
• AAMC (C, MC, MD)	• Avera (MC)	• CHIIC (MC)
• ACCF (C)	• BMC (MD)	• Clifford Beers (C, MD)
• Altarum (MD)	• CCNC (MD)	• Icahn (MC, MD)
• CCC (MC)	• FPHNY (MC, MD)	• NACHRI site 1 (MD)
• CHS (MD)	• Montefiore (C, MC, MD)	• NACHRI site 2 (MD)
• Columbia (MD)	• Northwell (MC)	• NACHRI site 3 (MD)
• DMC (C, MC, MD)	• SCH (MD)	• NHCHC site 1 (MD)
• FSCL (MC)	• U KS (MC)	• NHCHC site 2 (MD)
• Hopkins (MC, MD)	• U NC (C, MC)	• NHCHC site 3 (MD)
• Mesa (C, MD)	• UCSD (C, MC, MD)	• NHCHC site 4 (MD)
• NM (C, MC, MD)	• UIC (MD)	• NHCHC site 5 (MD)
• UHCMC model 1 (MC)	• UMich (C, MC)	• NYC H+H model 1 (MD)
• Ventura (MC, MD)	• UNM (MC, MD)	• UCSF model 1 (MC)
• VillageCare (C, MC, MD)	• Wash U (MD)	• UCSF model 2 (MC)
• Yale (MD)	• WI DHS (MD)	• UHCMC model 2 (MC)
	• NYC H+H model 2 (C, MC)	

Note: Amerigroup is not included in the table because the service delivery model will be supported by a state grant after the cooperative agreement ends, and the awardee did not propose another payment model.

C = commercial insurer; MC = Medicare; MD = Medicaid.

Awardees most frequently included Medicaid as the ultimate payer (24 awardees, 32 models), followed by Medicare (22 awardees, 24 models) and commercial payers (12 awardees, 12 models). Sixteen awardees included multiple payers by varying the components of their models. For example, Montefiore offers one payment model but is negotiating different rates and parameters of that model with MCOs for their Medicaid managed care, Medicare Advantage, and commercial plans.

D. Factors associated with progress in developing alternative payment models

In this section, we identify the three awardee characteristics most commonly associated with successful development of payment models. We then describe the four most commonly reported facilitators and the two most commonly cited barriers associated with their progress.

1. Awardees' characteristics associated with progress

The eight awardees (representing 15 models) that entered into a contract with at least one payer (Stage 3) were more likely to have (1) a model that included Medicaid beneficiaries, (2) prior experience with a similar program, and (3) a program spanning larger geographic regions, especially across multiple states. These trends were driven by a cluster of awardees that had two or more of these characteristics. Several of these awardees were able to advance multiple

payment models to the point of securing contractual agreements (Catholic Health Initiatives, National Association of Children's Hospitals, National Health Care for the Homeless, New York City Health + Hospitals).

Medicaid beneficiaries. One-third of the models that included Medicaid beneficiaries had a contract in place near the end of the third year of the cooperative agreement (11 of 32 models), compared with one-fourth of the models that included Medicare beneficiaries (5 of 24 models) and only one of 12 models that included beneficiaries who had commercial insurance (Table V.2). Many of the models that included Medicaid beneficiaries and for which the awardees had advanced to contractual agreements were from two multi-site, multi-state awardees who were able to build on pre-existing payment arrangements or strong relationships with payers (three site-specific models from National Association of Children's Hospitals and five site-specific models from National Health Care for the Homeless). For example, four of the sites for National Health Care for the Homeless were able to build on pre-existing contracts with Medicaid MCOs and Medicaid FFS.

Prior experience. The awardees were more likely to have progressed toward contractual agreements when they had significant experience with the same or a similar health care delivery model prior to the cooperative agreement (11 of 24 models). Only 4 of the 16 models developed by awardees that had less experience but had pilot tested the model had contractual agreements. In addition, only 3 of the 9 models developed by awardees that had no direct experience with a similar program had contractual agreements. We also found that awardees that had experience with similar models tended to have pre-existing payment arrangements or strong relationships with payers. These awardees include the same two multi-site, multi-state awardees (National Association of Children's Hospitals and National Health Care for the Homeless) plus three other awardees that were developing one model each (Catholic Health Initiatives, Mount Sinai, and New York City Health + Hospitals).

Larger geographic regions. Models developed by awardees operating in multiple states were more likely to have advanced to contractual agreements (11 of 15 models) compared to models developed by awardees operating solely within a single state (0 of 4 models), within a region within a state (2 of 8 models), or locally (3 of 19 models). Four awardees accounted for the 11 models from multi-state programs (Catholic Health Initiatives, National Association of Children's Hospitals and Related Institutions, National Health Care for the Homeless, and University of California at San Francisco).

2. Common facilitators

The most frequently cited facilitators of payment model development included (1) consulting with internal and external experts, (2) leveraging existing or upcoming alternative payment model initiatives, (3) early and continuous collaboration with payers, and (4) learning from peer organizations.

Consulting with experts. Twenty-four awardees reported that consulting with various internal and external payment model experts was a key support to develop their payment models. The experts included external actuarial consultants and health economists from affiliated

institutions. At least seven awardees cited CMMI and the HCIA R2 implementation and monitoring contractor as major sources of support.

Leveraging alternative payment model initiatives. To support their own model development efforts, many awardees leveraged existing alternative payment model initiatives and plan to leverage upcoming ones, including the Medicare Shared Savings Program, the Medicare Next Generation ACO model, the Medicare Oncology Care Model, the Medicare Skilled Nursing Facility Value-Based Purchasing Program, Medicaid Health Homes programs, and Medicaid ACOs and capitated risk contracts. Several awardees were able to or planned to integrate their programs into these initiatives in order to address challenges related to access to data and analytics, small target populations, and alignment with state policy initiatives.

Collaborating with payers. Eight awardees described early and continuous collaboration with payers as a key support. These awardees developed relationships with payers to share data and to continually revise the payment models, an approach that was beneficial to both the awardee and the payers. For example, Seattle Children's Hospital worked with four Medicaid MCOs and was able to get all four payers to agree to a shared payment model framework for children with medical complexity.

Learning from peer organizations. Six awardees also cited learning from peer organizations as a facilitator of model development. For example, the National Association of Children's Hospitals made payment model development an explicit part of the learning collaborative for its 10 sites, and leaders from the sites mentioned the importance of learning from each other. The University of California at San Diego learned from the deliberate approach taken by an HCIA Round 1 awardee to develop relationships with payers and to executing contracts after the end of its cooperative agreement.

3. Common barriers

The two most frequently cited barriers to payment model development were (1) acquiring data to assess operational costs and impacts on participant spending, and (2) the short duration of the programs.

Acquiring data to assess costs and outcomes. At least 13 awardees said that acquiring data was a major barrier to the development of their payment models. The challenges they faced included difficulty negotiating access to payer claims data, lack of data on potential comparison groups, and limited or no data on health care spending. For example, leaders from Boston Medical Center described how several Medicaid MCOs initially expressed interest in their program but ultimately did not share their data because of concerns about revealing differences in contracted rates to competing organizations. The MCOs that did share data did so only for program participants; no data were provided on other children that could have been used to compare outcomes with the participants.

Short program duration. Eleven awardees cited the short time frame of the cooperative agreement as a significant challenge. Awardee leaders noted that the combination of program start-up time, the expected exposure time needed for impacts to show up, and data lags meant that only two years of data or less were available for analysis prior to end of the third year of the cooperative agreement. Many awardees planned to continue analyses to support the development of payment models during their no-cost extensions or after the cooperative agreement ended.

> "Looking at savings over time takes time, and a three-year award period is really two years of data. To base a payment model on two years of data is tough."
>
> —Program leader

Fewer awardees noted three other barriers: (1) small target population, (2) concurrent state policy initiatives, and (3) lack of precedent for APMs. Six awardees said that it was a challenge to develop an alternative payment model for a relatively small target population. Awardee leaders noted that the small number of participants, whether in the program overall or in individual sites that would contract for a payment model, made it not only difficult to calculate reliable utilization and spending estimates but also limited the payers' interest in negotiating a specific payment model for a small group of beneficiaries. Six awardees also noted how broader state policy initiatives over-rode their efforts to develop a payment model. For example, the National Association of Children's Hospitals, the University of Illinois, and Washington University faced delays in negotiations because of the implementation or expansion of Medicaid managed care plan for their target populations; this factor required the awardees to switch gears from negotiating with a single state Medicaid agency to negotiating with multiple Medicaid MCOs and to wait for the Medicaid agency's managed care bidding and contracting cycle to be completed. Finally, six awardees cited the lack of precedent for non-FFS payment models for their programs' services or provider types as a major barrier. For example, Altarum and Columbia University noted that dental providers had little familiarity with APMs and had few incentives to participate. Community Care of North Carolina faced a similar challenge with pharmacists.

4. Association between payment model development and implementation effectiveness

We found no clear association between the awardees' progress in developing and implementing a payment model and implementation effectiveness. The lack of an association between payment model development and implementation effectiveness is not surprising, as program implementation and payment model development were often happening in parallel but were led by different staff with different skill sets. In addition, effective program implementation does not necessarily result in positive program impacts (such as improved health outcomes and lower costs), which we hypothesize would give awardees stronger leverage for negotiating a payment model. The evidence of program impacts by the end of the third year of the cooperative agreements was scant.

E. Implications for HCIA R2 goals and future similar programs

While most awardees have invested a significant amount of time and resources in developing payment models, relatively few have actually entered into a contract with any payers. As of August 2017, eight of the 38 awardees had entered into at least one contract with a payer.

These awardees varied in size from large, multi-site collaboratives (for example, National Health Care for the Homeless) to smaller, single-site organizations (Clifford Beers). The eight awardees also varied in their target populations from Medicare-enrolled adults receiving care in rural hospitals (Catholic Health Initiatives) to Medicaid-enrolled children with complex conditions (National Association of Children's Hospitals). The models varied substantially in the type and degree of financial incentives and risk. Thus, a substantial amount of work remains to be done if these models are to be implemented after HCIA R2 funding ends.

A majority of the payment models we analyzed included Medicaid as a payer (32 of 46 models), and awardees who took this approach generally made more progress in reaching a contractual agreement. This development was driven by a handful of awardees who successfully leveraged their experience with similar models, had ongoing discussions with payers about APMs, and shared their consulting resources and lessons from peer learning across multiple implementation sites. Despite these efforts, the awardees still faced significant challenges with acquiring and using data to support the development of their models, and many models were based on short-term analyses, the results of which were not robust enough to inform the longer-term sustainability of the programs under these forms of payment. Awardees that included Medicare as a payer had less success in advancing their models, and those that did relied primarily on existing Medicare APMs.

This page has been left blank for double-sided copying.

VI. SUSTAINABILITY, SCALABILITY, AND REPLICABILITY OF THE HCIA R2 PROGRAMS

A. Introduction and summary of findings

Awardees varied widely in their plans to continue their programs beyond the cooperative agreement (sustainability), to expand them to reach additional participants (scalability), and to encourage other organizations to replicate the programs (replicability). In the third program year, most awardees had made progress on these sustainability, scalability, and replicability (SSR) plans, which they initiated in the second program year. In this chapter, we explain our methods for assessing SSR, summarize the development of the awardees' SSR plans, describe the awardees' SSR progress in the third program year, explain the factors affecting SSR, and consider the next steps for SSR. Although our analysis describes the range of readiness, strategies pursued, and factors affecting SSR as reported by the awardees and their partners, it was difficult to assess both the likelihood of funding and whether SSR would happen as planned for many awardees. Awardees that received a no-cost extension commonly planned to use that time to continue their search for funding and conduct other SSR activities. Also, because impact analyses will be available for only a subset of the awardees and not until 2019, we cannot link the awardees' SSR progress during the cooperative agreement to our objective estimates of their success in reducing their enrollees' expenditures or improving the quality of care—the factors necessary to secure ongoing financial support from payers and other stakeholders. We will continue to track SSR progress, changes, and patterns across awardees throughout the remaining evaluation period.

The key findings on SSR from the end of Year 3 are summarized below.

- The extent to which awardees expected to have sufficient funding to sustain their entire programs varied widely by the end of the third program year. Seven awardees seemed ready and able to sustain their programs largely intact at most implementing sites. Sustainability was uncertain or unlikely for six awardees, and it varied a lot for the remaining 25 awardees.
- Sixteen awardees had scaled their programs, and seven awardees were in various stages of attempting to do the same. Other organizations replicated the programs of two awardees, and two additional programs were in the process of being replicated.
- Awardees reported that the following factors affected sustainability planning: (1) buy-in from staff and organizational leaders, (2) ability to demonstrate program's value in order to engage payers or other funders, (3) federal and state policies that support or align with the program, (4) alignment between the program and the implementing organization's mission and goals, and (5) a program design that is adaptable and flexible enough to overcome challenges to sustainability.
- We found no clear association between the awardees' plans to sustain, scale, or replicate their programs and overall implementation effectiveness. However, awardees that were effective in staff recruitment and training were more likely than other awardees to have advanced sustainability plans.

B. Methods for assessing SSR

As required by the terms of the HCIA2 initiative, awardees developed plans to sustain their programs after the end of the cooperative agreements, and some opted to expand their programs. This continuation and expansion occurred through three related but distinct activities: sustainment, scaling, and replication. We defined these terms at the end of the second program year to capture not only the extent to which the programs might endure beyond the cooperative agreement but also the extent to which the programs might be spreading to new areas to serve more people, both as extensions of the awardees' original efforts and efforts by other entities that have decided to adopt and implement them.

- **Sustainability** is the continuation of the program components, in whole or part, across the original provider organizations and target populations.
- **Scalability** is the expansion of the program to other populations, providers, sites of care, geographic areas, or payers. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- **Replicability** is the adoption of the program, in whole or in part, by another organization, not affiliated with the awardee, that serves individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardees' SSR progress (Section D) and the factors affecting it (Section E) was the implementation interviews we conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included open-ended questions about progress in sustaining, scaling, and replicating the programs and provided awardees with our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter (June to August 2017) for updated information about the awardee's SSR activities. Before discussing the awardees' progress, we summarize their baseline SSR plans, which we assessed from the 2016 implementation interviews and earlier awardee self-reports submitted to the implementation and monitoring contractor.

We provide an indication of the prevalence of each type of SSR planning and activity, and factors affecting SSR based on available data. Because our findings on SSR were drawn primarily from interview responses to open-ended questions about SSR, the respondents may not have thought of, been aware of, or had time to report on all relevant activities and plans during their interviews. As a result, our findings may not comprehensively cover awardees' SSR plans and experiences.

C. Development of SSR plans

In the second program year, we started tracking the degree to which the awardees were planning for sustainability, the types of strategies they were pursuing, and the factors affecting their planning and activities.

As documented in our second annual report, most (31) of the 38 awardees had developed sustainability plans and were pursuing ways to implement them. Yet just one-third of the awardees (13) had started to implement the strategies to help sustain their programs (in whole or part) beyond the cooperative agreement. These strategies typically revolved around financial solvency, including securing ongoing payments for their programs through public or private payers and improving program efficiency to reduce the cost of ongoing operations. Other strategies focused on developing the infrastructure to support the programs beyond the cooperative agreement (for example, health IT and staff expertise). These 13 awardees were more likely than the other awardees to have (1) had prior experience in implementing the program (or something similar), (2) met their enrollment targets, and (3) used passive enrollment.

In contrast to the 31 awardees that were actively working toward sustainability, seven of them had only minimally developed a sustainability plan as of the second program year. These seven awardees had to focus on more immediate program enrollment and implementation issues.

Although CMS did not require the awardees to develop plans to scale or replicate their programs, we detected an interest and activity in both as of the second program year. In general, scaling and replicating followed effective implementation and the establishment of a foundation for sustaining the program. In some cases, however, scaling and replicating were strategies to generate revenues to help sustain the core program. Three awardees had scaled all or part of their programs by the end of the second program year: Montefiore Medical Center, Cleveland Medical Center, and Avera. One awardee, Community Care of North Carolina, had replicated its program in multiple states, and two cities (Los Angeles and Anaheim) had replicated Mesa Fire and Medical Department's program.

D. SSR plans: Progress and changes in year three

In the third program year, the awardees mostly proceeded along the path to SSR that they started out on in the second year. The awardees that were farthest along in both planning for sustainability and in implementing these plans in the second program year typically tended to be those that were in a better position to sustain their programs by the end of Year 3. Awardees that did not have concrete sustainability plans as of the second program year were developing them in the third year, although two awardees (Nebraska Medicine and Yale University) were still in the early stages of the process. By the end of the third program year, the awardees differed widely in SSR readiness and activities. We discuss this variation in sustainability, scalability, and replicability in the following sections.

1. Sustainability

All 38 awardees that were operating at the end of the third program year were trying to sustain their programs after the cooperative agreement ended. No awardee expressed the intent to close down its program. The interest in moving forward stemmed from the awardees' perception that their programs were meeting—or at least approaching—their stated goals related to lowering costs or improving outcomes, even if they did not yet have robust information to confirm success.

Despite this enthusiasm, however, sustainability readiness varied considerably across awardees. This variation largely reflected the amount of funding and other resources obtained by the awardees relative to the costs of operating their programs. This relationship between resources and costs had implications for whether and how long an awardee could sustain its whole program. Thirty of the awardees received no-cost extensions ranging from 3 to 12 months; many said that they planned to use the no-cost extension to continue planning for sustainability. Below we discuss the overall sustainability readiness of the awardees, the importance of different types of funding and other resources, the awardees' efforts to streamline their programs, and how much of their programs they thought they could sustain.

a. Variation in readiness to sustain their programs

Table VI.1 shows the variation in awardees' readiness to sustain their programs, in whole or part. Seven awardees seemed to have the necessary plans and resources in place to continue their programs largely intact right after the cooperative agreement ended and into the foreseeable future (Table VI.1, category A, column 3). Such plans did not mean that sustainability would be guaranteed or without its challenges, and the seven awardees expressed concern that their resources might not ultimately be sufficient for years to come. For instance, Catholic Health Initiatives has a three-year contract with an ACO, but the awardee could not predict whether it would be renewed.

Seven awardees were poised to sustain their programs largely in full right after the cooperative agreement:

- Amerigroup
- Avera
- Catholic Health Initiatives
- Four Seasons
- National Healthcare for the Homeless
- New York City Health and Hospitals
- University of Kansas

At the other end of the spectrum, sustainability was rather uncertain or unlikely for six awardees. (Table VI.1, category D). Four of them—Clifford Beers Guidance Clinic, Johns Hopkins University, Northwell, and VillageCare—had been planning for sustainability but were struggling to implement a payment model or other funding support. These four awardees expressed considerable concern about their ability to secure enough sustainability funding during their 3- to 6-month no-cost extensions. Nebraska Medicine and Yale University were still in the early stages of planning for sustainability but expected to use their 6- and 12-month no-cost extensions, respectively, to develop stronger strategies for sustaining their programs.

The prospects for sustainability among the remaining 25 awardees are more varied and complex. These awardees expected to be able to sustain at least parts of their programs but were at different stages of acquiring the funding necessary to do so. Eight awardees of the 25 had funding to sustain parts of their programs (Table VI.1, category A, column 2). Another eight awardees were in negotiations with payers but were concerned about a gap in funding before their payment models start (Table VI.1, category B). Nine of the 25 awardees had the organizational commitment and/or other interim funding to sustain their programs (in whole or part) while pursuing or negotiating their payment model and other ongoing funding (Table VI.1, category C). Nine of the 25 awardees spanning the different categories had relatively long no-cost extensions (9 to 12 months), which might be enough time for the outlook on sustainability to improve.

Table VI.1. Awardees' sustainability readiness at the end of the third program year

Categories of funding secured or expected	Partial program	Whole program
A. Payment model and/or other ongoing funding secured or expected by end of cooperative agreement	Altarum CCC DMC NACHRI UHCMC UMich Ventura UCSF	Amerigroup Avera CHIIC FSCL NHCHC NYC H+H U KS
B. Active negotiations with payers but might face gap in funding before payment model starts	BMC SCH	CCNC Montefiore UCSD UIC UNM Wash U
C. Organizational commitment and/or other interim funding secured while pursuing ongoing funding (payment model or other)	ACCF CHS Columbia Icahn Mesa U NC	AAMC FPHNY WI DHS
D. Insufficient funding	Clifford Beers, Hopkins, NM, Northwell, VillageCare, Yale University,	

Although we assessed the prospects for sustainability for each awardee and its program overall, the awardees noted that the outlook might vary across their implementing sites. Awardees with multiple implementing sites typically and gradually delegated the responsibility for sustainability planning to the sites as they became capable of integrating the program into their organizations. For example, Catholic Health Initiatives reported that gradually transferring the responsibility for program costs to its implementing sites prepared the sites to fully operate the program after the cooperative agreement ended. Even so, awardee leaders typically planned to continue providing guidance to their sites. For instance, the National Association of Children's Hospitals required each of its 10 sites to develop a detailed sustainability plan; the awardee would then use the no-cost extension to help the sites implement their plans.

b. Reliance on both short-term and long-term funding

The awardees reported that they need both short-term limited funding and longer-term ongoing funding in order to cover the operating costs of their programs immediately after the end of the cooperative agreement and into the future. Most awardees were relying on their payment models with Medicaid, Medicare, and commercial payers to generate substantial and ongoing sources of funding. As discussed in Chapter V, all but one awardee had developed at least one payment model, but they were in different phases of implementing them. Eight awardees had a finalized contract with at least one payer, 15 were in negotiations with payers, and 14 were still developing their models.

Despite the strong relationship between payment model development and sustainability, the models were not always necessary to support sustainability, nor were they usually sufficient to do so, at least initially. Table VI.2 describes progress in payment model development and the prospects for sustainability for awardees that had the most developed payment models and sustainment plans by the end of the third program year (the seven awardees discussed above, Table VI.1, category A, column 3). Only three awardees had payment model contracts in place and were ready to sustain their programs in their entirety: Catholic Health Initiatives, National Healthcare for the Homeless, and New York City Health + Hospitals. Five awardees had a contract in place, but they were not enough to ensure sustainability. For example, two awardees had payment models in place that would allow them to scale, rather than sustain, their programs (Clifford Beers Guidance Clinic and the University of California at San Francisco). Three awardees had payment model contracts that supported just partial sustainability: the National Association of Children's Hospitals, Mount Sinai, and Cleveland Medical Center.

Other ongoing funding was supplementing or even replacing the payment models. Four awardees secured enough financial support from other sources to sustain their programs largely in full into the foreseeable future: Amerigroup, Avera, Four Seasons, and the University of Kansas. This support came from the following sources: (1) the state or their own organizations, (2) insurance reimbursement through existing billing codes (particularly new Medicare billing codes for chronic care and transitional care management), (3) integrating the program into an existing ACO, and (4) fees from the implementing sites. Avera, Four Seasons, and the University of Kansas also continued to pursue not only payment models for additional and longer-term support but also more value-based payment approaches (see Chapter V).

Table VI.2. Relationship between progress toward a payment model and sustainability readiness

Awardees with payment model contract in place but full sustainability not ensured	Awardees with payment model contract in place and ready to sustain whole program	Awardees without payment model contract but have other funding to sustain program
Clifford Beers. Has payment model in place to scale modified program to commercial patients; does not have Medicaid payment model to sustain initial program	**CHIIC.** Shared savings program ACO contract and implementing sites are covering costs of sustaining program	**Amerigroup.** Partner received state funding to continue program and did not propose a separate payment model
Icahn. Has one payment model contract and is developing payment models and payment processes for additional insurance companies; does not yet have funding to sustain program in full but has interim funding to sustain program immediately after cooperative agreement ends	**NHCHC.** Program continuing at all implementing sites through a Medicaid payment model and additional funding sources	**Avera.** Retail subscription model fees will sustain program at most original sites (33 of 45) and at all 20 scaled sites while pursuing payment model
NACHRI. Four of 10 sites have payment model in place to sustain program; remaining six sites still developing proposals or negotiations; only partial sustainment (in terms of sites and program components) seems likely, at least initially	**NYC H+H.** Incorporating program into existing Medicare and Medicaid global risk contracts, and into New York's Delivery System Reform Incentive Payment (DSRIP) program	**Four Seasons.** Implementing sites, Medicare billing codes, and hospice revenue will cover costs of sustaining program at most sites (4 of 5) while pursuing payment model
UCSF. Has Next Gen ACO contract, but this is a pilot to scale the program to determine whether it will work under a capitated payment model; a federal grant and new Medicare billing codes will allow partial sustainability		**U KS.** Program will be sustained at all sites through revenues from existing ACO and patient safety organization, Medicare billing codes, and hospital funds while pursuing payment model
UHCMC. Payment model is CMS Oncology Care Model, which will cover aspects of the program for a portion of participants		

ACO = accountable care organization.

The awardees also relied on short-term and supplemental funding from public sources (federal, state, and local) and private sources (for example, their own organizations and private foundations). These funds would (1) help to sustain the programs temporarily between the end of the cooperative agreement and a fully implemented payment model, (2) provide more time to build the case for the payment model, or (3) help cover remaining costs. For example, the University of California at San Francisco received a grant from the National Institute on Aging to cover about half the program's costs for five years and to continue collecting data to demonstrate the program's values to potential payers. Eight awardees were concerned they would not have sufficient short-term funding and other resources to continue their programs while awaiting ongoing payments (Category C of Table VI.1). For example, although Boston

Medical Center was confident that its payment model would ultimately be approved, one of its two sites was concerned that a gap in funding between the end of the cooperative agreement and the anticipated start of the payment model could threaten sustainability.

c. Streamlining programs to cut costs

Awardees reported that they streamlined program operations as a way to reduce the costs of their programs and make it easier to sustain them. They also wanted to present their programs as efficient and therefore more appealing to potential funders. Common streamlining efforts included cutting or reducing minor program components and integrating components into existing health IT and other infrastructure and workflows. For example, Ventura County Health Care integrated its care guidelines into its EMR so that providers could access them beyond the cooperative agreement. Most of CareChoice's sites were ending the transition coordinator role and moving its functions to social workers or other existing staff. In addition, the Fund for Public Health in New York was monitoring participants' needs and the demand for care coordinator visits in order to understand when it would be appropriate to reduce the frequency of such visits. Strategies reported less often included removing an administrative "layer" and cutting back on monitoring the staff's compliance with the program. For example, Amerigroup and the University of Michigan transferred program leadership to partner organizations, and Community Care of North Carolina reported that it relaxed its monitoring efforts because program functions had become second nature for staff.

d. Whole versus partial sustainment

Sixteen awardees planned to sustain their entire programs (after streamlining and typically with minor modifications) at all or most of their original sites and for their original participants (third column in Table VI.2). However, another group of 16 awardees acquired enough funding, at least initially, to sustain their programs only partially (first column in Table VI.2). These awardees made this decision based on the total funding available to them or on what their current funders wanted to support. Partial sustainment typically meant continuing only select program components, target groups or sites. For instance, Mesa Fire and Medical Department was sustaining about half of its program (the 911 response function through mobile community medicine units) but not the other (the hospital transition component). Columbia University ended the community health worker position, the core part of its program, but was able to maintain the education piece for dental residency students. In a few cases, awardees predicted they would have enough resources to serve just a subset of participants. For example, the University of Michigan expected to have a payment model in place to sustain the program for just its commercially insured population, and Detroit Medical Center expected to have funding to continue the program at just two of its three clinics. These awardees and others like them typically hoped to eventually obtain more funding to reinstate the components that they put on hold or to reach their full target population.

For the six awardees that were behind in or struggling with their sustainability plans (Category D of Table VI.1), we did not have enough information to categorize how much of their programs, if anything at all, they would sustain.

2. Scalability

The awardees' efforts to scale their programs in the third program year were much greater than they were in the second year. Scaling was typically motivated by the awardees' desire to serve more participants, which would also save money through economies of scale and, in a few cases, generate additional revenue for the program. The awardees usually scaled or planned to scale their programs even if they did not have complete information about the program's impact. They scaled either by expanding the program to new populations, spreading the program to other providers in their system or community, or working with additional payers to obtain reimbursement for additional participants.

Sixteen awardees scaled their programs:

- Association of American Medical Colleges
- Altarum
- Avera
- Catholic Health Initiatives
- Children's Health Society
- Cleveland Medical Center
- Four Seasons
- Icahn School of Medicine at Mount Sinai
- Montefiore
- National Association of Children's Hospitals
- National Health Care for the Homeless
- University of Kansas
- University of North Carolina
- University of New Mexico
- Washington University
- Wisconsin Department of Health Services

By our definition of scaling (i.e., expanding the program to other populations, providers, sites of care, geographic areas, or payers, but not as a means to meeting enrollment targets for the cooperative agreement), 16 awardees had scaled their programs by the end of the third program year. Most of these awardees scaled essentially their whole programs and made minor modifications. For example, the University of Kansas and the University of New Mexico were implementing the program at new provider sites by the end of the third program year. One of the two implementing hospitals participating in the Wisconsin Department of Health Services' program expanded eligibility for the program. Three of the National Association of Children's Hospitals' sites were implementing their program's care processes across all of their affiliated primary care and outpatient complex care practices, having found that it was easier for staff to complete the processes when they were part of the standard of care for everyone.

Three of the 16 awardees only partially scaled their programs. For example, Altarum scaled its SmileConnect website, which started in Michigan, to 30 states, but its two other main program components either were in the process of being scaled or did not have plans for scaling (the dental registry and training/technical assistance, respectively).

Seven other awardees were in the process of scaling their program or key components of it. For instance, Amerigroup was scaling its whole program by lowering the age of eligibility and including youth from the juvenile detention system. The University of Michigan was scaling a streamlined version of its program to all surgical patients and to pre-operative pain planning as well.

The need for sustainability funding prompted scaling in six cases. Avera, Clifford Beers Guidance Clinic, and Johns Hopkins University reported scaling their programs to help generate

revenue through patient referrals or other fees to sustain their programs. Four awardees (Catholic Health Initiatives, Montefiore Medical Center, New York City Health and Hospitals, and Boston Medical Center) reported scaling their programs to align with their payment models—particularly to add providers to match the provider networks in their new ACO arrangements.

The remaining 16 awardees were evenly split between those that were in early discussions about scaling and those that did not report any intention of scaling their programs.

3. Replicability

In contrast to scaling, program replication remained rare. Only two awardees reported that their programs had been replicated by others. Community Care of North Carolina reported that 15 other pharmacy networks had replicated its program, and another 25 or more had requested technical assistance for replicating the program. The awardee planned to obtain clinically integrated network status in order to charge pharmacies a fee for technical assistance, which would provide funds to help support sustainability. The replicated versions of Mesa Fire and Medical Department's programs held steady in Anaheim and Los Angeles, yet the awardee was struggling to gain support from payers to sustain its original program. Payers may have been reluctant to support the program in Mesa because they found it to be more viable in larger jurisdictions.

Two awardees replicated their programs:

- Community Care of North Carolina
- Mesa Fire and Medical Department

Two awardees were helping other organizations replicate their programs:

- Altarum
- Four Seasons

Seven awardees were in the process of scaling:

- Amerigroup
- Boston Medical Center
- Clifford Beers
- Johns Hopkins
- University of California San Francisco
- University of Illinois
- University of Michigan

Despite these issues, the awardees' replication efforts increased by the end of the third program year, and two awardees were actively helping other organizations replicate their programs. Altarum, which operates its program in Michigan, was working with the University of California, Los Angeles to replicate the program in Los Angeles County. The awardee also planned to discuss the possibility of replicating its program with two other states that expressed interest. Four Seasons was finalizing an agreement to serve as a consultant to another organization within the same county that was interested in replicating the awardee's program. Although Washington University was not explicitly attempting to replicate its program, its payment model, if approved, would provide an incentive for providers across Missouri to provide the program's services to Medicaid enrollees.

Thirteen awardees were attempting to foster replication by disseminating information about their programs. Common strategies included presenting at conferences, writing journal articles about their programs, meeting with other interested organizations and/or creating toolkits for them. For example, Mount Sinai received a grant from a foundation to develop a program implementation manual containing step-by-step guidance for others interested in replicating the

program. In a unique strategy, the entity that is sustaining Amerigroup's program, Families First, trademarked COACHES, the program name, and certain components to maintain program integrity by controlling the circumstances under which another organization could use the program's name. However, replication activities did not proceed as planned for two awardees. CareChoice's early discussions and plans to create a toolkit stalled while the awardee focused on its payment model. And the University of Michigan's original plan to sell its risk assessment tool to other health systems did not come to fruition.

Twenty awardees did not report any significant efforts or plans to replicate their programs. Still, a few of these awardees reported that other states, providers, or payers had expressed interest.

E. Factors associated with progress in implementing SSR plans

Awardees often identified the same challenges and facilitators associated with SSR. In this section we show which factors were most commonly reported by the awardees, and how they differ for awardees grouped by how likely they were to sustain their programs. Also presented are the findings from our analysis of the relationship between the awardees' SSR progress and implementation effectiveness.

1. Facilitators and barriers to SSR

The facilitators and barriers to SSR reported by the awardees fall into five categories that relate to their abilities to acquire ongoing funding and other resources to support their programs: (1) demonstrating program value, (2) engaging internal stakeholders, (3) engaging external stakeholders, (4) aligning their programs with state and federal policy, and (5) program design. The first four factors are similar to those we reported in the second annual evaluation report; program changes emerged as a new factor in the third year.

Demonstrating program value. The awardees' ability to show that their programs were meeting their goals was key to gaining funding and other support for SSR. Although awardees lacked robust data on their programs' impacts, many were able to demonstrate program value in other ways. For example, the Association of American Medical Colleges reported that that the organization's leaders saw that providers liked the program and therefore supported its sustainment despite the awardee's inability to confirm a positive return on investment. To demonstrate the value of its program to organizational leaders and payers, the University of Michigan program used participants' testimonials about the importance of the program and internal assessments that the program showed promise in saving money and improving patient outcomes.

On the other hand, almost one-third of the awardees reported that insufficient and untimely data on program impacts continued to present substantial barriers to SSR, especially with respect to gaining support from payers. The awardees commonly reported that the three-year cooperative agreement did not provide adequate time to (1) achieve the intended effects on participant outcomes, especially in light of implementation delays the awardees experienced; (2) obtain data, which may have runout periods; or (3) finish rigorous analyses of program data necessary to engage payers.

Engaging internal stakeholders. Awardees also noted the importance of engaging both internal leaders and frontline staff to be able to continue and potentially expand their programs. For example, program leaders at the National Association of Children's Hospitals reported that hospital leaders not only fostered a positive shift in attitudes of clinical and administrative staff regarding the program but also committed existing research and health IT staff to continue many of the program's routine processes after the cooperative agreement.

In many cases, maintaining support from internal leaders and staff was challenging. Four awardees reported concerns about maintaining internal support for their programs because of leadership changes or because of the potential need to ask more of implementing sites and staff. For instance, the Wisconsin Department of Health Services thought that hospital leaders' commitment to the program could wane if future funding support was low and if the hospitals had to absorb more costs over time. Many awardees faced having to rebuild the buy-in and expertise of program staff because they had lost many of their original staff who were insecure about the stability of their jobs after the cooperative agreement.

Engaging external stakeholders. Engaging external stakeholders was vital to obtaining significant ongoing financial support. These stakeholders commonly included state legislators and state agencies (particularly Medicaid), public and private insurers, and private foundations. As discussed above in Section B and in Chapter V, awardees were at different stages in these relationships with external stakeholders. For example, the Clifford Beers Guidance Clinic had secured one new contract with a commercial payer and hoped that the payer's large market presence and strong reputation would prompt other payer's interest in the program. On the other hand, Yale University reported plans to start engaging payers during or after the no-cost extension.

Alignment with state and federal policy environment. State and federal policies continued to facilitate or hinder SSR depending on whether the incentives aligned with program objectives. State and federal policies that promoted value-based care particularly helped SSR for awardees that shared that focus. For instance, Community Care of North Carolina reported that the state Medicaid program's interest in payment reform helped to garner support for the program and prompted potential replication by others. New state regulations or policies prompted or facilitated scaling for two awardees: the Association of American Medical College's program because of a Virginia state mandate to make a particular pediatric specialty expertise more available throughout the state, and Altarum's program because of a new Michigan policy that required all dentists participating in Medicaid to use the program's dental registry.

A third of awardees reported that federal and state policies—particularly related to funding and payment—could hinder SSR. For example, Washington University was concerned about the reliability of Title X funding, which is a critical source of support for its program, and Catholic Health Initiatives lamented that its rural health clinic sites would not directly benefit from the MACRA payment adjustments that helped its other sites sustain the program. A common concern was that Medicare or Medicaid do not reimburse for new staff types or services. For example, The University of California San Diego found telephone interactions to be effective ways to treat low-income patients, but struggled to continue them because they typically are not billable under Medicare and Medicaid. Columbia University expected that state Medicaid

programs' reluctance to pay for community health workers' efforts to prevent early childhood caries would impede program replication.

Program design. Program design also affected SSR in a few ways. First, program features and objectives that matched organizational objectives facilitated SSR. For example, leaders at the University of Kansas and Catholic Health Initiatives reportedly committed to sustaining the programs because they were aligned with the organizations' focus on population health. Second, programs that were not overly customized to the awardee or that could be easily modified helped SSR. For example, New York City Health + Hospitals and Columbia University refined their programs to be "turn-key," meaning the program required minimal training and provided protocols and technology to support implementation by others.

Program changes. Although most awardees thought their streamlining and minor program modifications would not harm, and would potentially improve, their prospects for SSR, it is not yet clear whether changes such as cutting staff positions and integrating their roles into the roles of existing staff will be viable. Also, awardees that expected to be able to sustain, scale, and replicate their programs only partially could face more challenges with SSR. Five such awardees expressed concern that fragmented programs could result in a significant loss of program fidelity and potentially hinder SSR. For example, Altarum received state funding to maintain the health IT component of its program but not for training, and it quickly observed a significant drop in clinician certifications to use that component. A few awardees acknowledged they did not know what the ultimate effects of cutting back certain components would be. As one program leader said, "You can have a tripod and take one of the legs off, but which direction it's going to fall, you never know."

> "You can have a tripod and take one of the legs off, but which direction it's going to fall, you never know."
>
> —Program leader

We also examined the common barriers and facilitators to SSR reported by awardees that were most and least likely to sustain. For the seven awardees poised to sustain their programs, common facilitators included early and ongoing sustainability planning, gaining buy-in from internal leaders, and program alignment with organizational mission and priorities as well as with federal and state policy, especially value-based payment approaches. Yet these awardees still grappled with high program costs and lack of proof of program effectiveness as they look ahead to long-term sustainability. The six awardees least likely to sustain their programs were held back by their inability to demonstrate program value and/or misalignment with federal and state policy, particularly an inability to bill for program services and use staff in new ways.

2. Association between SSR and program design and implementation

We detected little meaningful association between SSR and program design and implementation. Three of the seven programs that were in a good position to be sustained targeted high-risk adults with chronic conditions; the remaining four had different target populations. The set of six awardees unlikely to sustain their programs targeted a range of populations. We detected no patterns in sustainability either by the main payers targeted by awardees (Medicare, Medicaid, etc.) or by the type of program services they provided. Care

coordination and care management were the most common types of program services among the awardees most and least likely to sustain their programs, but not disproportionately so. Finally, for overall implementation effectiveness, we looked for patterns of progress toward sustainability across the domains of implementation effectiveness discussed in Chapter III (for example, progress toward Year-3 enrollment projections and patient engagement). We found no clear association between the awardees' plans to sustain, scale, or replicate their programs and overall implementation effectiveness. However, awardees that were effective in staff recruitment and training were more likely than other awardees to have advanced sustainability plans.

F. Looking ahead

While our SSR analysis described the range of readiness, strategies pursued, and factors affecting SSR as reported by the awardees and their partners, we faced challenges assessing for many awardees the likelihood of the necessary steps and funding actually coming through. Also, while the no-cost extensions provided awardees more time to work on SSR and potentially make notable progress, they also increased the challenge of assessing what the awardees would actually accomplish. Lastly, final impact analyses will not be completed until August 2019, and even then only for 23 of the awardees. Thus, we cannot link awardees' SSR progress during the cooperative agreement to our objective estimates of their success in reducing enrollees' expenditures or improving the quality of care—the factors likely necessary to secure financial support from payers and other stakeholders for SSR. We will continue to track SSR progress, changes, and patterns across awardees throughout the remaining evaluation period.

VII. NEXT STEPS

A. Implementation evaluation

Now that the three years of the initial cooperative agreements have ended, the primary focus of the implementation evaluation will shift from collecting and analyzing evidence of implementation effectiveness to aligning the findings from the impact analysis with awardees' implementation experience. The implementation evaluation will also help us to identify the best approach for analyzing outcomes for the 15 programs for which a full difference-in-differences impact analysis will not be possible.

We will also complete the last of our qualitative data collection and analysis activities during the next six months. This includes (1) reviewing any final progress reports and supplemental materials (such as enrollment reports and self-monitoring and measurement reports) that the awardees submit to the implementation and monitoring contractor at the end of their agreement, and (2) conducting follow-up interviews with the awardees that received a no-cost extension about the continued progress on their service delivery and payment reform models.

Although 30 awardees received a no-cost extension for anywhere from 3 to 12 months, only six awardees planned to continue enrolling beneficiaries, though most will be providing services to existing enrollees during some or all of this period. We will follow up with these awardees to assess whether additional data are available that provide evidence of the programs' implementation effectiveness. In addition, we will also consider conducting follow-up interviews with awardees that have promising payment models or sustainability, scalability, or replicability plans. Finally, we will analyze the data from the patient survey to assess the participants' experience with the programs and their perceptions of the care they received under the intervention.

The remaining implementation evaluation activities and their timing during the final two years of the HCIA R2 evaluation are shown in Table VII.1. We will document the findings from these efforts in the final individual awardee narratives and synthesize the findings in the final report, due in late 2019.

Table VII.1. Remaining implementation evaluation activities

Activity	Purpose	Schedule
1. Review final awardee documents	Complete our final assessment of implementation effectiveness and the factors associated with program success	February–August 2018
2. Conduct follow-up interviews with awardees that received a no-cost extension	Update our assessment of awardees' implementation experience with program enrollment and delivery of services during the no-cost extension	February–August 2018
3. Conduct follow-up interviews with awardees that have promising payment models or sustainability plans		February–August 2018

Table VII.1 *(continued)*

Activity	Purpose	Schedule
4. Analyze data from patient survey	Assess participants' experience and satisfaction with program, as well as their perception of the effect of the service delivery model on access to care, quality of care, and health outcomes	February–August 2018
5. Integrate qualitative and quantitative findings	Use findings from implementation evaluation to interpret and corroborate results from impact evaluation	Ongoing as results become available
6. Help identify approaches to evaluating outcomes for programs not receiving impact analysis	Develop alternative methods for assessing changes in health outcomes and expenditures that could be potential effect of the program	Ongoing as needed

B. Patient survey

We launched the patient survey (the third and final survey conducted for this evaluation) in May 2017 and concluded data collection in October 2017. The survey included participants from 21 of the 38 awardees. As with the other two completed surveys (staff and clinician), some of the content was customized to the characteristics of each awardee's program, but all of the surveys included questions about the patients' use of program services, their experience and satisfaction with those services, and their opinions about the effect of the program on both the quality of care and their health.

Our target response rate was 35 percent, with the goal of a minimum of 300 eligible completes per awardee for awardees with more than 1,000 eligible participants. We selected a random sample of approximately 1,000 participants for awardees with 1,000 or more eligible participants, and we sent surveys to a census of participants for awardees with fewer than 1,000 eligible participants. Although the definition of survey eligibility varied by awardee, it was generally based on when respondents last participated in the program (to optimize patient recall) and on their insurance type (we focused on Medicare, Medicaid, and CHIP beneficiaries).

We received 2,940 returns across all 21 awardees, with individual awardee response rates ranging from 5.6 percent for the University of Illinois at Chicago to 55.8 percent for Johns Hopkins University (Table VII.2). We are currently coding the written responses and preparing summary tables of all the survey questions by awardee. We will include a summary of the results of the patient survey in the final quarterly report (due in May 2018) and in the final evaluation report (due in late 2019).

Table VII.2. Patient survey response rates, by awardee

Awardee	Number of sampled participants	Number of completed surveys	Response rate[a] (%)
Avera Health	1,009	118	28.2
Boston Medical Center	215	50	23.3
Community Care of North Carolina	1,009	151	23.3
Catholic Health Initiatives Iowa Corp.	601	119	44.3
City of Mesa Fire and Medical Department	1,184	92	11.9
Detroit Medical Center, Vanguard Health Systems	1,009	104	12.4
Fund for Public Health in New York, Inc.	1,009	148	18.7
Johns Hopkins University	330	164	55.8
Icahn School of Medicine at Mount Sinai	424	107	31.4
Montefiore Medical Center	1,009	107	11.7
Nebraska Medical Center	279	135	51.3
New York City Health and Hospitals Corporation	1,009	74	8.5
Seattle Children's Hospital	432	62	15.7
Regents of the University of California San Diego	1,009	292	37.8
The Board of Trustees of the University of Illinois – A	926	42	5.8
The Board of Trustees of the University of Illinois – B	1,009	44	5.6
University of Kansas Hospital Authority	1,009	295	39.9
Village Center for Care	891	277	33.8
Ventura County Health Care Agency	1,009	110	18.7
Washington University School of Medicine in St. Louis	1,009	163	16.8
Wisconsin Department of Health Services	472	150	32.0
Yale University	457	136	37.2

[a]We used the American Association for Public Opinion Research's (AAPOR) response rate calculation methodology No. 4 to measure the awardee response rates. The AAPOR methodology is defined as: Completes / [completes + eligible nonrespondents + eligibility rate*(nonrespondents with unknown eligibility)], where the eligibility rate is (completes + eligible nonrespondents) / (all cases with known eligibility).

C. Impact evaluation

The conclusions from the impact evaluations in this report are preliminary because the treatment groups consist only of the beneficiaries who enrolled in a program during roughly the first two years of the cooperative agreement and outcomes were measured during only the first year after enrollment. The estimates therefore cover only the early period of program operations, when the awardees were still learning how best to implement their interventions and focusing substantial resources on identifying and enrolling patients, and on engaging both providers and patients in the interventions. Patients also had a relatively short exposure to programs services over this time frame.

In the remainder of this evaluation contract, we will conduct propensity-score matching that includes all Medicare and Medicaid beneficiaries who enrolled over the entire period in which the awardees' programs were operated, for the 23 awardees for which credible impact estimates

are possible. With the larger samples and longer follow-up periods, we will estimate models to generate final impact estimates for the life of each program and for different intervals of follow-up time. We will also perform more extensive sensitivity and robustness checks. In addition, we will estimate program impacts in a Bayesian framework. This approach will allow us to improve the precision of our impact estimates for each program by borrowing strength across estimates across multiple time periods and outcomes. The result will be more precise estimates for any given time period or outcome. The Bayesian modeling will also allow us to estimate the probability that a program impact exceeds a specified value that we deem to be substantively meaningful. We will also identify and implement strategies for evaluating programs for which we will not be able to conduct a rigorous impact analysis using awardee-provided and claims data for participants, if possible

Throughout the remaining months before the final report is due to CMS toward the end of 2019, we will continue to provide informal memoranda to CMS showing results from matching and impact analyses for individual awardees as they become available.

APPENDIX A

Staff and Clinician Surveys

This page has been left blank for double-sided copying.

Table A.1. Number of completed surveys and response rates for staff and clinician surveys, by awardee and in total

	Staff survey		Clinician survey	
Awardee	Number of completed surveys	Response rate (%)	Number of completed surveys	Response rate (%)
AAMC	--	--	129	49.4
ACCF	24	83.3	111	56.9
Altarum[a]	11	100.0	167	61.9
Altarum[a]	40	62.5	–	–
Amerigroup	24	82.8	–	–
Avera	42	72.4	–	–
BMC	20	95.2	–	–
CCC	57	78.1	–	–
CCNC	6	100.0	–	–
CHIIC	50	96.2	107	64.1
CHS	6	66.7	–	–
Clifford Beers	27	90.0	4	100.0
Columbia	16	80.0	58	81.7
DMC	20	69.0	–	–
FPHNY	23	79.3	–	–
FSCL	25	100.0	22	91.7
Hopkins	21	100.0	–	–
Icahn	5	83.3	24	66.7
Mesa	18	94.7	10	90.9
Montefiore	23	82.1	69	71.1
NHCHC	7	50.0	–	–
NM	17	73.9	–	–
Northwell	10	100.0	–	–
NYC H+H	33	71.7	–	–
SCH	8	88.9	10	100.0
U KS	148	85.1	32	88.9
U NC	--	--	74	67.3
UCSD	18	95.0	–	–
UCSF	24	96.0	–	–
UHCMC	8	72.7	–	–
UIC	36	78.3	–	–
UMich	32	76.2	81	69.8
UNM	34	53.1	4	44.4
Ventura	12	92.3	72	52.2
VillageCare	10	90.9	–	–
Wash U	13	100.0	10	100.0
WI DHS	24	96.0	18	100.0
Yale	22	50.0	–	–
Total	914	79.7	1,002	63.3

Note: A '–' indicates that awardee did not receive a survey. The response rate was calculated as the number of completes divided by the number in the sample minus ineligible cases. Because the individual awardee survey samples were small, we did not weight them to adjust for nonresponse. A description of the survey methodology and survey instruments is available from the authors on request.

[a]Altarum received two staff surveys: one for program staff at Altarum and one for non-clinical staff at the implementing sites who received training through the intervention. Clinicians at those sites completed the clinician survey as well.

This page has been left blank for double-sided copying.

APPENDIX B

MATCHING RESULTS

This page has been left blank for double-sided copying.

A. Background

Nearly all of the comparison groups used as counterfactuals for the impact analysis will be selected by using the method of propensity score matching (PSM). The purpose of PSM is to allow us to construct a comparison group that is similar to the treatment group on key observable characteristics that affect treatment status and outcomes. The method uses logistic regression to estimate the probability that each treatment and potential comparison group beneficiary are actually in the treatment group. Each treatment group beneficiary is then matched to one or more beneficiaries in the potential comparison pool with a similar probability of being in the treatment group. The covariates used in the logistic model are referred to as "matching variables." When the distribution of each matching variable is similar in the treatment and comparison groups, the groups are said to exhibit good "covariate balance," and the match is typically considered acceptable.

This appendix presents matching results for three awardees: CareChoice, Nebraska Medicine, and New York City Health + Hospitals.[16] It also compares results for two randomized control trials: the University of California at San Francisco and the University of Illinois at Chicago. We used three measures to assess the quality of the match.

- **Standardized differences.** The standardized difference is defined as the treatment-comparison difference between the mean values of a covariate, expressed in standard-deviation units. Smaller standardized differences indicate more closely matched groups. A commonly invoked benchmark (Rubin 2001) suggests that groups are well matched if standardized differences for all covariates are less than 0.25. However, we strove for differences no larger than 0.10.
- **Equivalence tests.** An equivalence test expresses the null hypothesis as stating that the absolute value of the difference between two means is greater than a specified amount. The tests we conducted specified a difference in covariate means of at least 0.25 standard deviations. Rejection of the null suggests a good match.
- **T-test for difference in means.** We also conducted a standard t-test for differences in the mean value of each covariate. Unlike the two measures above, the t-test is not typically recommended as a test for the quality of a match because samples that are large enough will often lead to rejection of the null hypothesis. In many cases, the differences are statistically significant but operationally insignificant. Even more concerning, when samples are small, the test can fail to reject the hypothesis of equal means for the two groups even when the differences are large because the power of the test is low. We included this test because it could signal issues with the match that deserve further investigation. Note that the t-test and the equivalence test can both reject the null hypothesis (that is, the means are not equal for

[16] The status of the matching process for all awardees is shown in Table B.3 at the end of this appendix. Two awardees—the University of California at San Francisco and the University of Illinois—used randomized designs, so matching was unnecessary. Appendix A in these awardees' narratives includes mean values for covariates for these two awardees and for the three awardees in which PSM was used.

the two groups, but the difference does not exceed 0.25 standard deviations), especially when samples are large.

B. Matching results

By the measures just introduced, the overall quality of matches for the three awardees is excellent. Detailed tables showing covariate balance and results of each test appear in Appendix A of these awardees' narratives. In nearly every instance, standardized differences are far less than 0.25, and equivalence tests indicate that differences in covariate means are less than 0.25 standard deviations. Despite the good match, we must stress that a matched comparison group design cannot provide the protection against bias that results from unobserved variables in the way that randomized control groups do.

Table B.1 summarizes the covariate balance for the three awardees as expressed by three representative variables: Medicare spending prior to enrollment, HCC score, and dual-enrolled status. In every case, the standardized difference in means is less than 0.1, and the equivalence test strongly rejects the hypothesis that the difference exceeds 0.25 standard deviations. In one instance—prior Medicare spending at New York City Health + Hospitals—the treatment and comparison groups differ by more than 10 percent, and the treatment/comparison means are significantly different from one another at the 0.05 level. As noted above, this result does not itself indicate that the match is compromised, especially given the large samples for the two awardees.

Table B.1. Measures of covariate balance for three representative variables: propensity score matching awardees

Awardee	Treatment	Comparison	Percent difference	Standardized difference	Difference *p*-value	Equivalence *p*-value
CCC (901/2,957)						
Medicare spending[a]	$2,564	$2,614	-2.0	-0.02	0.63	0.00
HCC score	2.48	2.48	-0.1	-0.00	0.98	0.00
Proportion dual enrolled	0.12	0.13	-7.1	-0.03	0.54	0.00
Nebraska Medicine (351/1,394)						
Medicare spending[b]	$8,474	$8,087.	4.6	0.03	0.72	0.00
HCC score	2.01	2.04	-1.3	-0.02	0.80	0.00
Proportion dual enrolled	0.31	0.31	0.0	0.00	1.00	0.00
NYC H+H (2,042/6,145)						
Medicare spending[a]	$2,120	$1,869	11.9	0.06	0.04	0.00
HCC score	1.52	1.52	0.6	0.01	0.80	0.00
Proportion dual enrolled	0.54	0.54	0.9	0.01	0.75	0.00

Notes: Treatment and comparison group sample sizes (n_T/n_C) appear in parentheses after the awardees' names.

The standardized difference is calculated as the difference in means divided by the treatment group standard deviation. Stated *p*-values result from a weighted two-sample t-test. The equivalence test comprises two one-sided tests of whether the difference between the treatment and comparison group means exceeds 0.25 standard deviations (in either direction) of that variable. If we reject that hypothesis, we are confident that the true difference between the two populations being compared is less than 0.25 standard deviations.

[a] Medicare spending PBPM in the 12 months prior to program enrollment.

[b] Medicare spending PBPM in the 3 months prior to program enrollment.

By way of comparison, Table B.2 shows the covariate balance for the University of California at San Francisco and for the University of Illinois, the two awardees that implemented

randomized designs. The variables for the University of Illinois differ from those presented for other awardees because the intervention targeted Medicaid beneficiaries.

Table B.2. Measures of covariate balance for three representative variables: randomized control trial awardees

Awardee	Treatment	Control	Percent difference	Standardized difference	Difference *p*-value	Equivalence p-value
UCSF (359/180)						
Medicare spending[a]	$999	$1,067	-6.8	-0.06	0.55	0.03
HCC score	1.32	1.38	-4.2	-0.06	0.55	0.03
Proportion dual enrolled	0.14	0.11	20.2	0.08	0.36	0.03
UIC (821/6,819)						
Medicaid spending[a]	$372	$391	-5.1	-0.01	0.81	0.00
CDPS score	3.64	3.28	9.8	0.08	0.07	0.00
Number of outpatient ED visits for asthma	0.18	0.15	20.95	0.05	0.27	0.00

Notes: The standardized difference is calculated as the difference in means divided by the treatment group standard deviation. Stated p-values result from a weighted two-sample t-test. The equivalence test comprises two one-sided tests of whether the difference between the treatment and comparison group means exceeds 0.25 standard deviations (in either direction) of that variable. If we reject that hypothesis, we are confident that the true difference between the two populations being compared is less than 0.25 standard deviations.

The size of the treatment and comparison group samples (nT/nC) appear in parentheses after the awardees' names.

[a] Medicare spending PBPM in the 12 months prior to program enrollment.

[b] Medicaid spending PBPM in the 12 months prior to program enrollment.

CDPS = Chronic Illness and Disability Payment System; ED = emergency department; HCC = hierarchical condition category.

C. Further considerations

In our view, the matching results shown in Appendix A of these awardees' narratives support the use of these three comparison groups to estimate impacts of the corresponding interventions on beneficiary outcomes. Nonetheless, the matching effort has furnished some lessons and elements that deserve special attention for the remaining matching efforts.

PSM models should be accompanied by judgment. The PSM model is an algorithm that generates the closest possible match between treatment and comparison group beneficiaries on included characteristics; the propensity score is used as a gauge. The observed covariate balance is itself a by-product of the algorithm's operation, which makes no distinction among covariates except through their effects on the propensity score. Viewed from the perspective of existing research and literature, this situation is not ideal. All covariates are not equal. Prior spending is known to be strongly correlated with future spending, and therefore, the 12 percent difference in prior spending for treatment and comparison groups at New York City Health + Hospitals might be viewed as somewhat too great a difference for a proper comparison group. That difference can probably be reduced by using a caliper, resulting in a greater imbalance for some other variable or variables. Whether this tradeoff is acceptable is a matter of research judgment, which we will exercise to a greater extent as matching continues.

Issues in matching are sometimes not immediately apparent. A comparison group that appears well matched may be discovered to be ill-suited to the impact analysis even though all

measures suggest good covariate balance. For example, we had considered the matching process complete for one awardee—all covariates were exceptionally well matched. It was only when we developed preliminary impact estimates that we became aware that despite the excellent match, the trajectory of utilization and expenditures within the 12 months prior to enrollment/pseudo-enrollment were quite different for the treatment and selected comparison groups. For obvious reasons, we have re-started the matching process for this awardee to better track the pre-enrollment pattern of expenditures over time for the two groups. Because we can never anticipate all of the issues or problems that might arise as the analysis continues, we expect that we may need to revise other comparison groups if and when unexpected results appear.

Matching plans. Over the coming months, we plan to continue the matching process with additional awardees. We will continue to provide CMS with matching memos describing the quality of the matches and their suitability for impact estimation as the final or interim matching is completed.

Table B.3. Status of matched comparison groups

Awardee	Status of matched comparison group
AAMC	Matching delayed due to need for SK&A data; now in process
ACCF	Matching delayed due to need for SK&A data; now in process
Altarum	Recently received Medicaid data for awardee, but matching delayed due to difficulties in matching at provider level; anticipate completing matching and sending memo to CMMI prior to final report
Amerigroup	No rigorous evaluation
Avera	Matching on full sample is about to take place
BMC	No rigorous evaluation
CCC	Matching complete; memo sent to CMMI
CCNC	Matching in process
CHIIC	In the process of drawing a comparison group
CHS	No rigorous evaluation
Clifford Beers	Medicaid data not yet available
Columbia	Medicaid data not yet available for large enough sample
DMC	No rigorous evaluation
FPHNY	Matching in process; will send memo shortly after AR3
FSCL	No rigorous evaluation
Hopkins	No rigorous evaluation
Icahn acute	No rigorous evaluation
Icahn rehab	No rigorous evaluation
Mesa	In the process of drawing the comparison group
MMC	Medicaid data not yet available for large enough sample
NACHRI	Medicaid data not yet available for large enough sample
NHCNC	No rigorous evaluation
NM	Matching complete; memo sent to CMMI
Northwell	No rigorous evaluation
NYC H+H	Matching complete; memo sent to CMMI

Table B.3 *(continued)*

Awardee	Status of matched comparison group
SCH	Medicaid data not yet available
UCSD	No rigorous evaluation
UCSF	Randomized control group; balance memo sent to CMMI
UHCMC	Matching revisions in process
UIC	Randomized control group; balance memo sent to CMMI
UKS	Matching delayed as inconsistencies between finders file and Medicare claims were resolved; now in process
UMich	No rigorous evaluation
UNC	No rigorous evaluation
UNM	No rigorous evaluation
VillageCare	Matching started but then delayed after finding that a substantial percentage of participants received services on the date of enrollment; will complete matching after AR3
Ventura	Stage of COPD is an important piece of clinical data that allows us to differentiate between those with COPD and those at-risk of developing COPD. It is likely that we will need to restrict our treatment group to those with COPD so that we can find similar comparison beneficiaries with a diagnosis of COPD in Medicare claims data. The data we have received to date have been very incomplete. We are attempting for one last time to get the most complete information on staging before drawing the comparison group.
Wash U	No rigorous evaluation
WI DHS	Medicaid data not yet available for large enough sample
Yale	No rigorous evaluation

This page has been left blank for double-sided copying.

APPENDIX C

IMPACT METHODOLOGY

This page has been left blank for double-sided copying.

A. Overview

Estimating program effects requires several steps: Identification of the treatment group, choosing a method for estimating the counterfactual, building the analytic samples and variables, and estimating the statistical models. In this appendix, we describe each of these steps. Perhaps the most important component is estimating the counterfactual—that is, the outcomes that would have been expected for enrollees had the program not existed. We used two different approaches to estimate this counterfactual. For the two awardees who were willing and able to randomize potential participants, we used a randomized control trial (RCT). For the other three awardees, we used a difference-in-differences longitudinal panel design.

B. Identifying the treatment and comparison or control analytic samples

To evaluate the impact of the program, we follow treatment and comparison or control beneficiaries over time. For programs enrolling beneficiaries with Medicare fee-for-service (FFS) coverage, we defined the treatment group as individuals who enrolled in the program, met the awardee's eligibility criteria at the time of enrollment, and met basic criteria needed to ensure that we would have the data needed for the analysis. These. basic criteria are that, on the date of enrollment, an individual must be (1) alive, (2) enrolled in Medicare Parts A and B, (3) have Medicare as the primary payer, and (4) be in Medicare FFS. The individual must have satisfied these criteria for at least 90 days in the baseline period, defined as the 365 days preceding the enrollment date. For the one awardee for whom we evaluated program impacts on a Medicaid population, beneficiaries had to be enrolled in Illinois Medicaid, either FFS or managed care, during at least one month of the analysis period. Some awardees also require beneficiaries to have a qualifying event as a condition of enrollment (for example, treatment in an emergency department [ED] for an ambulatory care sensitive condition). For RCTs, the treatment and control beneficiaries are determined at the time of randomization.

We defined comparison beneficiaries as individuals who did not enroll in a program throughout the entire program period; who met all of the awardees' eligibility criteria that could be assessed with Medicare/Medicaid claims and enrollment files; and who met the four data availability criteria mentioned above. Because no enrollment date exists for individuals who were not in the treatment group, we defined for each potential comparison group beneficiary a "pseudo-enrollment" date that we used to define the end of the baseline period and beginning of the follow-up period. The method for setting the pseudo enrollment dates varied across awardees, depending on how the treatment group established eligibility (e.g., the date of a qualifying event, such as a hospital discharge or ED visit).

1. Propensity score matching methods

For the awardees that are not conducting RCTs, we constructed a comparison group by using propensity score matching methods. The matching is intended to minimize nonrandom selection bias in our estimates of the program impacts by producing a matched comparison group that is similar on average to the treatment group on key observable baseline covariates. For most awardees, we matched at the individual beneficiary level. If, however, a potential comparison beneficiary could have had multiple qualifying events throughout the program period, we matched at the event-level before selecting the qualifying event that resulted in the closest match

to the treatment beneficiary. This means that in each approach, there is only one pseudo-enrollment date per potential comparison beneficiary.

Data. We used Medicare claims and enrollment data from 2013 to 2016 to identify the treatment and comparison beneficiaries. For propensity score matching, we constructed key covariates by using beneficiary information for the baseline period, which is 12 months prior to the enrollment date (or the pseudo-enrollment date for comparison beneficiaries). Because enrollment occurred from 2014 to 2016, we used administrative data starting in 2013 to define the baseline period.

Methods. We used logistic regression models to estimate the probability that the individual is in the treatment group on the pooled sample of treatment group members and potential comparisons, then used the estimated equation to predict propensity scores for each of the individuals in the sample. We examined distributions of propensity scores by treatment and potential comparison group for all quarters in the baseline period. We calculated the propensity score distance between each treatment beneficiary and each potential comparison. The propensity score summarizes all of the matching variables, giving more weight to variables that more strongly predict treatment status so that the distances capture how similar each potential comparison beneficiary is to each treatment beneficiary.

Propensity score matching entails forming matched sets of treatment and comparison beneficiaries who share a similar value of the propensity score. We applied optimal matching methods to select those comparison beneficiaries from the potential comparison pool who, in the aggregate, were most similar to treatment beneficiaries on the matching covariates.[17] Optimal matching minimizes a global distance criterion instead of many local criteria as used by nearest neighbor matching. In our approach, optimal matching minimizes the total sum of differences on the estimated propensity scores between the treatment and matched comparison beneficiaries. To implement optimal matching, we used the optmatch package available at the Comprehensive R Archive Network (CRAN). The *fullmatch* function within this package creates optimal full matches for the specified treatment group.[18] It was our standard practice to apply a caliper to the propensity score between matched beneficiaries, which is a limit on the maximum acceptable propensity score distance between beneficiaries in a matched set.

Across awardees, we matched with replacement, whereby each potential comparison beneficiary could be matched to more than one treatment beneficiary. Using optimal matching methods with replacement, we can put similar treatment beneficiaries into the same matched set—allowing the same comparison beneficiary to match to more than one treatment beneficiary—if doing so will reduce the overall propensity score distance. In most cases, we used a default matching ratio, allowing any given comparison beneficiary to match to up to five treatment beneficiaries and up to five comparison beneficiaries to match to any single treatment

[17] Rosenbaum, P.R. "A Characterization of Optimal Designs for Observational Studies." *Journal of the Royal Statistical Society, Series B,* vol. 53, no. 3, 1991, pp. 597–610.

[18] Hansen, B.B., and S. Klopfer. "Optimal Full Matching and Related Designs via Network Flows." *Journal of Computational and Graphical Statistics*, vol. 15, 2006, pp. 608–627.

group beneficiary. The actual ratio for any matched set depends on how many beneficiaries fall within the caliper and satisfy the various restrictions imposed, such as exact matching on critical variables. This optimal matching approach makes it easier to achieve good overall balance. The restrictions on the number of treatment group cases to which a single comparison case can be matched helps limit the size of the design effect due to weighting, since the matching weights applied to comparison beneficiaries in our analysis are equal to the inverse of the matching rate.

Matching variables. We included a number of beneficiary characteristics as predictors in the propensity score model, such as demographic characteristics, prior (baseline) expenditures and utilization, and risk scores, and indicators of various chronic conditions. The exact list of matching covariates varied by awardee. We exact-matched on a subset of variables deemed crucial to creating a credible comparison group. For these variables, treatment and comparison beneficiaries within any matched set must have the same value. For some awardees, we also applied calipers to a subset of variables that constrain the maximum difference between matched beneficiaries on the covariate. We did this because we wanted the caliper to help us achieve good balance on covariates that may strongly predict the outcomes of interest in the analysis.

Matching weights. We constructed matching weights to account for the number of matched comparison beneficiaries per treatment beneficiary so that the comparison beneficiaries matched to any given treatment group case have the same total weight as the comparison beneficiaries matched to any other treatment beneficiary. Each treatment beneficiary received a weight of 1, and each comparison beneficiary received a weight of $\frac{n_i^T}{n_i^C}$, where n_i^C is the number of comparison beneficiaries in a matched set *i* and n_i^T is the number of treatment beneficiaries in the same matched set. For example, if one treatment beneficiary is matched to four comparison beneficiaries who are not matched to other treatment beneficiaries, then each of the four comparison beneficiaries received a weight of 1/4. If, instead, these four comparison beneficiaries are all matched to two treatment beneficiaries, then each of the four comparison beneficiaries received a weight of 1/2.

Assessing the balance. After matching, we used several approaches to determine whether the treatment and comparison groups were adequately balanced, as described below. By balance, we mean that the two groups are very similar on mean values of all the covariates used in the matching model, and perhaps other characteristics as well. The results of these tests informed the quality of the match and alerted us to the possibility that we might need to modify the propensity score model used to select the comparison group.

As a first test, we calculated the standardized differences on all matching variables as well as additional key variables not included in the propensity score model. The standardized difference is a standard statistic to assess the balance of a sample before and after matching.

We calculated the standardized difference on a variable *X* as:

$$(1) \qquad \frac{\hat{\mu}_X^T - \hat{\mu}_X^C}{s_X^T}$$

In this expression, $\hat{\mu}_X^T$ represents the sample mean of variable X in the treatment group, $\hat{\mu}_X^C$ represents the weighted sample mean of the same variable in the selected comparison group, and s_X^T represents the standard deviation of the variable in the treatment group. Group-level means are calculated using individual beneficiary-level matching weights. We used the standard deviation of the variable in the treatment group rather than the pooled standard deviation in the denominator so that pre- and post-matching standardized differences are on the same scale.

The goal for matching is to have standardized differences less than 0.1 in absolute value for each matching variable. We strove to achieve this standard, though we recognize that we cannot always do so in practice, particularly when samples are small. We also calculated pre- and post-matching standardized differences on the propensity score, which summarizes all of the covariates and thus can serve as a measure of overall balance. Further, we checked balance on variables that we did not include in the matching model; our standards for balance on these variables are generally less stringent than our standards for matching variables.

In addition to looking at standardized differences, we performed additional balance tests to assess the quality of our matching. We report the omnibus test, which tests the joint hypothesis that the treatment and comparison groups are balanced (that is, the two groups' means are equal) across all matching variables.[19] The test statistic first measures the combined difference between the treatment and comparison groups across covariates and their linear combinations, and then assesses whether that difference is larger than we would expect to occur by chance, if they were drawn from the same population. We did not emphasize the results of this test because many small to moderate differences, well within the range that we consider equivalent for practical purposes, could lead us to reject the null hypothesis even when the groups are adequately balanced. To focus on differences that are meaningful in practice, we also conducted univariate equivalence tests, which test the null hypothesis that the difference between treatment and comparison group means exceeds an acceptable range. Rejection of this null hypothesis provides confidence that even if the two groups' means differ, the difference is small enough that it is unlikely to create biased estimates of program effects in our impact analysis. For these equivalence tests, our null hypothesis is that the difference exceeds 0.25 standard deviations, the value specified by statisticians as an acceptable level of difference. The test provides a conservative basis, therefore, for drawing conclusions about the similarity of the two groups, because, unlike traditional tests, small sample sizes do not increase the likelihood of concluding that the two groups are sufficiently similar. Comparing equivalence tests to traditional tests of the null hypothesis that the difference in means is equal to zero refines our understanding of the likely size of the difference.

[19] Hansen, B.B., and J. Bowers. "Covariate Balance in Simple, Stratified and Clustered Comparative Studies." *Statistical Science*, vol. 23, no. 2, 2008, pp. 219–236.

In addition to assessing standardized differences on all covariates and the closeness of the propensity scores of treatment group beneficiaries and their matched comparison beneficiaries, we also calculated the *p*-value from testing for whether the two groups' means are equal on these measures.

2. Random assignment of patients

Two awardees (the University of Illinois and the University of California at San Francisco) conducted RCTs in which beneficiaries were randomly assigned to either the treatment group or the control group. To assess whether random assignment resulted in two equivalent groups, we examined whether the treatment and control groups are statistically equivalent on pre-program, or baseline, characteristics as described more fully above. Overall, we found that the random assignment resulted in well balanced groups for both awardees (for more detail, see the individual awardee narratives in this report).

C. Estimation of program impacts

We examined the impact of the program on several key outcomes of interest that can be broadly grouped into three categories: (1) Medicare FFS expenditures, (2) Medicare service utilization, and (3) likelihood of acute care service utilization. For the one awardee included in this report that served primarily Medicaid enrollees, we estimated the impact of the program on a smaller set of outcomes that included Medicaid expenditures, service utilization, and likelihood of acute care utilization. We analyzed a common core set of outcome measures across awardees when applicable to the program (see Table C.1). We also estimated effects on awardee-specific outcome measures that are only applicable to a given program.

We examined outcomes in six-month periods from the beneficiary's enrollment date; however, the results presented in the body of this report are only for 12-month intervals, to reduce the variance in the outcome measures. Nonetheless, having results for six-month intervals enabled us to study whether program impacts vary with the length of beneficiaries' exposure to the intervention. At this stage of the evaluation, we have claims-based follow-up data on individuals for up to two years after they enter a program so that there can be as many as four semi-annual follow-up periods. Below we discuss the different approaches we used to estimate the impact of the awardees' programs, depending on whether we had a RCT or a difference-in-differences longitudinal panel design.

Table C.1. Outcome measures for Medicare expenditures, service utilization, and likelihood of service utilization

Variable
Medicare expenditures (per beneficiary per month)
Total
Acute inpatient
Other inpatient
Home health service
Outpatient
Skilled nursing facility

Table C.1 *(continued)*

Variable
Physician/supplier
Hospice
Durable medical equipment
Medicare service utilization (annualized per 1,000 beneficiaries)
Number of acute care hospitalizations beneficiary per year
Number of ED visits and observation stays that did not result in hospitalization
Number of primary care visits in ambulatory settings
Number of specialist visits in any setting
Number of specialist visits in ambulatory settings
Any health care utilization measures (i.e., likelihood of use measures)
Percentage with any acute care hospital hospitalization
Percentage with an ED visit or observation stay that did not result in hospitalization
Likelihood of a 30-day unplanned readmission among all beneficiaries

Note: All measures are weighted to reflect part-year Medicare FFS eligibility. Measures of expenditures are per beneficiary per month. We did not price-standardize the expenditure measures. Measures of service utilization are annualized and are per 1,000 beneficiaries. Measures of any health care utilization reflect the likelihood of use.

ED = emergency department.

1. RCT models

We observed individuals in the baseline period as well as in follow-up periods $t = \{1, 2, \ldots, P\}$, where t represents six-month periods. The maximum follow-up period for a program is *P*, but beneficiaries might not be observed for all periods if they died or lost FFS eligibility subsequent to their enrollment in the program. Because we use an intent-to-treat design, treatment group beneficiaries who dropped out of the program (or never engaged) are retained in the analysis for the entire evaluation period, even for those time periods after they dropped out.

In the RCT approach, the awardee randomly assigned beneficiaries into either a treatment group or a control group. We could obtain unbiased estimate of program impacts by computing simple differences in the mean values of outcomes between the treatment and control groups. However, we can obtain more precise impact estimates by including covariates and estimating regression models of the following form:

$$(2) \qquad y_{i,t} = \propto + \theta_t . treatment_i + \beta' . X_i + \sigma' . C_i + \vartheta' . M_i + \epsilon_{i,t}$$

Where $y_{i,t}$ is the outcome of individual *i* in period *t* (for example, total monthly Medicare expenditures during the *t*-th time period since he or she enrolled); $\propto$ is a constant term; $treatment_i$ is an indicator for whether the individual is randomized into the group that received program services; X_i are beneficiary characteristics such as gender, age, hierarchical condition category (HCC) scores, and other pre-enrollment characteristics, including baseline values of outcome measures; C_i are other characteristics that may affect outcomes such as community

features (for example, number of primary care physicians in geographic area) or hospital characteristics (for example, occupancy rate). M_i is a categorical variable indicating the maturity of the program in semi-annual periods relative to when the beneficiary was enrolled (or pseudo-enrolled, for comparison beneficiaries). For example, $M_i = 1$ if the beneficiary were enrolled during the first 6 months of program operations, $M_i = 2$ if the beneficiary were enrolled during the program's sixth to twelfth months of operation, and so on. For awardees with multiple facilities or treatment sites, we also included site indicators in equation (2) to account for potential differences in beneficiary outcomes by site. $\epsilon_{i,t}$ is a random disturbance term.[20]

We estimated equation (2) separately for each period, so we obtained an estimate of all parameters (the Greek letters in the equation), for all periods $t = \{1,2,\ldots,P\}$. The key parameter of interest is θ_t, which measures the impact of the program in participants' *t*-th period after enrolling. We can assess how program impacts vary with enrollees' length of exposure to the program because we estimated impacts for each of the semi-annual follow-up periods. We also estimated effects over a 12- and a 24-month period by constructing outcomes for these longer periods. These estimates for the single 12-month period are the only estimates reported in the body of the report.

2. Difference-in-differences models

For the awardees that did not conduct RCTs, we used a difference-in-differences framework to evaluate the programs, estimating the pre-post changes in outcomes for beneficiaries in the treatment group and for the matched comparison group over the same period. We measured effects relative to the baseline period—the 12 months before the beneficiary's enrollment date (or pseudo-enrollment date for the comparison group). The impact of the program was estimated as the difference between the average change over time for treatment beneficiaries and the average change over time for the matched comparison beneficiaries.

We follow the treatment and comparison groups over time, pooling beneficiaries with different enrollment dates and estimating impacts jointly for all 6-month intervals. Equation (3) specifies the regression model we used to estimate the impact of the program in the difference-in-differences framework:

$$(3) \quad Y_{i,t} = \alpha + \beta'.X_i + \sigma'.C_i + + \vartheta'.M_i + \delta_1 P_1.HCC_i + \delta_2 P_2.HCC_i + \ldots + \delta_P P_P.HCC_i + \tau.treatment_i + \gamma_1 P_1 + \gamma_2 P_2 + \ldots + \gamma_P P_P + \theta_1.treatment_i.P_1 +$$

[20] Note that, in this model, impacts are not allowed to vary with program maturity or site. In future analyses, we will interact both M_i and (if sample sizes permit) site binaries with treatment status to assess how impacts vary with these two factors.

$$\theta_2.treatment_i.P_2 + \ldots + \theta_P.treatment_i.P_P + \omega_{i,t}$$

The meaning of the notation is as described for equation (2) above. One key difference, however, is that for the difference-in-differences model, we estimate a single equation with separate observations for each individual for each time period, including the baseline period. This approach allows us to estimate differences in outcomes between baseline and each 6-month follow-up period, for both the treatment and comparison groups. Specifically, P_t is the period indicator, where $P_t = 1$ in follow-up period *t* and 0 otherwise. The other key difference is that we do not include baseline values of outcomes in the set of regressors, *X*, as we do for the RCT design, because those baseline values are included as separate observations in the difference-in-differences model. We included other individual beneficiary characteristics as before. However, we also included the interaction between the HCC score at baseline and the follow-up period $\left(P_t.HCC_i\right)$. We did this to reflect the expected different association between the fixed baseline HCC scores and the outcomes in early versus later follow-up periods. In cases where there are multiple treatment sites or facilities, we included site indicators to account for potential differences in beneficiary outcomes by site. In this model, α is a constant term and $\omega_{i,t}$ as a random error term.

As in our discussion of RCT models (Section C.1), the Greek letters are parameters to be estimated. The parameter τ estimates the treatment-comparison difference in an outcome during the reference period—the 12-month period before the beneficiary enrollment date; γ_t measures changes in the outcome for the comparison group over time; and θ_t's are the difference-in-differences estimates of the program impact.

The key parameters of interest—the θ_t's—represent the change in outcomes after the program enrollment date for the treatment group, net of the change in outcomes for the comparison group. Given that we included beneficiaries for up to *P* semi-annual periods, there are *P* post-enrollment impact estimates. In this model, we adjusted the standard errors for clustering at the unique beneficiary level to allow for serial correlation of the outcomes of individual beneficiaries over time in our longitudinal data set.

To construct impact estimates for the 12-month and 24-month periods, we computed equally weighted average of the parameters θ_t that contribute to the estimate for that timeframe; that is, for a longer timeframe comprised of *P* periods, our estimate is: $\sum_{t=1}^{P} \theta_t / P$. The 12-month impact estimates are thus the average of the impact estimates for the 1-6 and 7-12 month periods.

Again, our framework is an intent-to-treat approach because individuals are categorized as being in the treatment group or in the comparison group for the entire analysis period.

3. Estimation

We used ordinary least squares with an identity link function, in Stata, to estimate our models. For all outcomes, we calculated the impact estimates in each period in addition to the regression-adjusted means by treatment group status. In all analyses, we calculated robust standard errors (clustered at the unique beneficiary level in the longitudinal panel designs).

In the longitudinal panel approach, we used the matching weights described in an earlier section of this appendix to equalize the contributions of each matched set of comparisons per treatment beneficiary. We also used eligibility weights to account for the number of months the beneficiary was alive and in Medicare FFS from the start of the period to either the end of the period or the date of death. The final weight for the main difference-in-differences analysis was the product of the matching and the eligibility weights. In the RCT approach, we used only the eligibility weights in estimation.

Covariates. Although the RCT and our propensity score matching resulted in comparison groups that are very similar to the treatment groups along many dimensions, there may still be important differences that affect the outcomes under study. As such, we strove to capture such differences in the covariates of our model. We included three broad types of control variables in the vector X_i: (1) demographic characteristics, (2) measures of health status (such as HCC scores), and (3) chronic condition or historical diagnosis indicators. For some awardees, we also included community characteristics (for example, median income in the beneficiary's zip code) or characteristics of a medical care setting (such as number of hospital beds). These characteristics are included in the vector C_i above. As discussed, we also accounted for how long the program has been running relative to the beneficiary enrollment (or pseudo-enrollment) date, which is reflected in M_i. We included the HCC scores and follow-up period interactions to account for differences in sample composition over time. Finally, we included site indicators when the awardee had multiple treatment sites. The exact set of covariates we included in the impact estimation varies by awardee.

Construction of confidence intervals. We obtained the confidence intervals of the true program impacts by using different levels of significance. We produced the standard 95 percent confidence interval as well as the 90 and 99 percent confidence intervals. This allowed us to assess the variation in the intervals under different margins of error.

D. Sensitivity analyses

We verified the robustness of our impact estimates by conducting several sensitivity analyses where we modified the specification. We describe these analyses below. In subsequent analyses, we will conduct more sensitivity tests and tests of robustness of the impact estimates.

Extending the baseline period. We modified the baseline period to include the full two years before the enrollment date—as opposed to only one year as in the main impact estimates. We then estimated program impacts including the same covariates as in the main analysis. Our preferred specification is setting the 12 months before the program as the baseline period. In this

approach, impact estimates were measured relative to a reference period immediately preceding the enrollment date, which may better capture the health status of beneficiaries at the start of the program. However, this sensitivity analysis enabled us to understand how the choice of the baseline period affects our impact estimates. In general, we found that increasing the length of the baseline period had little effect on the impact estimates.

Top-coding outcomes. We studied the sensitivity of our results to top-coding outcome variables. Specifically, we accounted for outliers with extreme values by top-coding at the 99th percentile—that is, we replaced all values above the 99th percentile with the value of the outcome variable at the 99th percentile. We then estimated the model by using the top-coded variables. The findings from this analysis allowed us to understand how outliers may be affecting the estimated impact of the program.

APPENDIX D

INDIVIDUAL AWARDEE PROGRAM NARRATIVES

This page has been left blank for double-sided copying.

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: The Association of American Medical Colleges

April 18, 2018

Dana Jean-Baptiste (Mathematica Policy Research)
Ella Douglas-Durham (Mathematica Policy Research)
Brant Morefield (L&M Policy Research)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

The Association of American Medical Colleges used funding from HCIA R2 to implement the Coordinating Optimal Referral Experience (CORE) program (Table I.1). The goal of the

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The Association of American Medical Colleges stopped collecting data on new patients after August 31, 2017, and will track their use of all health services for up to six months, through February 28, 2018.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

CORE program was to reduce long wait times for specialty appointments and increase the effectiveness of referral processes by improving communication and care coordination between primary care providers (PCPs) and specialists. To accomplish the latter, the CORE program used two tools, eConsult and eReferral, collectively referred to as eCR.

An eConsult is an electronic tool that facilitates an exchange initiated by a PCP who is seeking clinical guidance from a specialist for a patient whose care the PCP wants to continue to manage. The eConsults were intended to avoid unnecessary in-person visits to specialists. If the consulting specialist thought the patient needed to be seen in person, however, he or she converted the eConsult to an eReferral for an in-person visit. The eReferral tool replaced the previous referral pathway used at participating sites. It was intended to enhance the referral process by preventing duplication of diagnostic testing and providing the specialists with comprehensive patient medical history. Both tools can either be used independently of each other or in sequence, if appropriate. The CORE program's innovation stems from the way it formalizes and streamlines communication between PCPs and specialists, leverages the electronic medical record (EMR) system to facilitate communication, pulls patient diagnostics and medical history data from the EMR, and expands access to feedback from a specialist.

The CORE program, which was launched on September 1, 2014, relied on health information technology (IT). Both the eConsult and eReferral platforms were embedded within the EMR system that was used by all academic medical centers (AMCs) participating in the CORE program. The participating AMCs included (1) Dartmouth-Hitchcock; (2) the University of California, San Diego; (3) the University of Iowa; (4) the University of Virginia; and (5) the University of Wisconsin. The primary focus of the CORE program was the PCPs employed by the AMCs at both AMC-based and community-based clinics. Site-level staff (for example, executive sponsors and PCP leads) introduced PCPs to the eConsult and eReferral tools at participating sites.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The Association of American Medical Colleges implemented the CORE program to enhance care delivery at the primary care–specialty care interface by giving PCPs decision support tools that allow them to seek guidance about patients' treatment and assess the appropriateness of referrals to specialist.
Major innovation	• Formalized communication between PCPs and specialists • Leveraged the EMR system to facilitate communication • Pulled patient data from the EMR • Expanded access to specialist feedback
Program components	Outpatient care coordination, decision support, health IT
Target population	• The primary focus of the CORE program was PCPs employed by the AMCs at both AMC-based and community-based clinics. • The target population included all patients older than 17, regardless of payer status, who visited the primary care practice sites.
Theory of change/ theory of action	Association of American Medical Colleges hypothesized that combining improved coordination and communication between PCPs and specialists with the eConsult interface would lead to a reduction in unnecessary subspecialty referrals and visits, more efficient use of specialist care, and improved access to specialists.

Table I.1 *(continued)*

Program characteristic	Description
Payment model	New fee-for-service (FFS) payment
Award amount	$7,125,770
Effective launch date	9/1/2014
Program setting	Primary care practices, hospitals, AMCs
Market area	Rural, urban, suburban
Market location	CA, DC, IL, IA, NH, VA, WI
Target outcomes	• Increased patient satisfaction • Decreased emergency department utilization • Decreased total cost of care • Decreased number of referrals • Increased quality of eConsults • Decreased cost for diagnostic testing and imaging • Increased eConsult use • Increased access to specialty care • Decreased out-of-pocket costs to patients (estimated)

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the Association of American Medical Colleges was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on four factors. First, the awardee enrolled 128,721 participants—107 percent of its enrollment target—by the end of the initial cooperative agreement. Second, the awardee was able to deliver services consistent with the original design and goals of the program. Third, the awardee was able to engage providers to use the tools. Finally, participating providers reported that the program had a positive effect on the delivery of care.

Impact evaluation. While we do plan to carry out a rigorous assessment of the impacts of the Association of American Medical Colleges' CORE program, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. The Association of American Medical Colleges did not develop a new payment model, instead pursuing options under existing fee-for-service (FFS) payment. The awardee provided a new service through the eConsult tool and identified different ways to reimburse for this service, including developing new billing codes and including the eConsults as a qualifying service within an alternative payment model (APM).

Sustainability plans. The Association of American Medical Colleges scaled the program to other member AMCs, affiliated providers, and patient populations. Although the Association of American Medical Colleges had plans for further scaling, it did not report plans to replicate the program to external, non-AMC sites or providers.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source was a survey of clinicians on their perceptions of the program's effect on care delivery. The survey was fielded from March to June of 2017 with a sample size of 300 clinicians and a response rate of 49 percent. The majority of respondents (81 percent) were PCPs. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

From a pool of AMCs that submitted proposals to participate in the CORE program, the Association of American Medical Colleges selected five AMCs based on the awardee's assessment of readiness, which included the AMCs' use of specific modules within their EMRs. All five AMCs participating in the program are members of the Association of American Medical Colleges. The CORE program focuses its intervention services on primary care clinics whose providers are employed by the five participating AMCs.

The CORE program does not provide services directly to patients. Instead, the program trains clinicians who treat the target population, and considers all patients served by PCPs at participating AMC sites as indirect participants. Indirect participants include all patients older than 17 years of age, regardless of payer status, who visit the primary care practice sites and

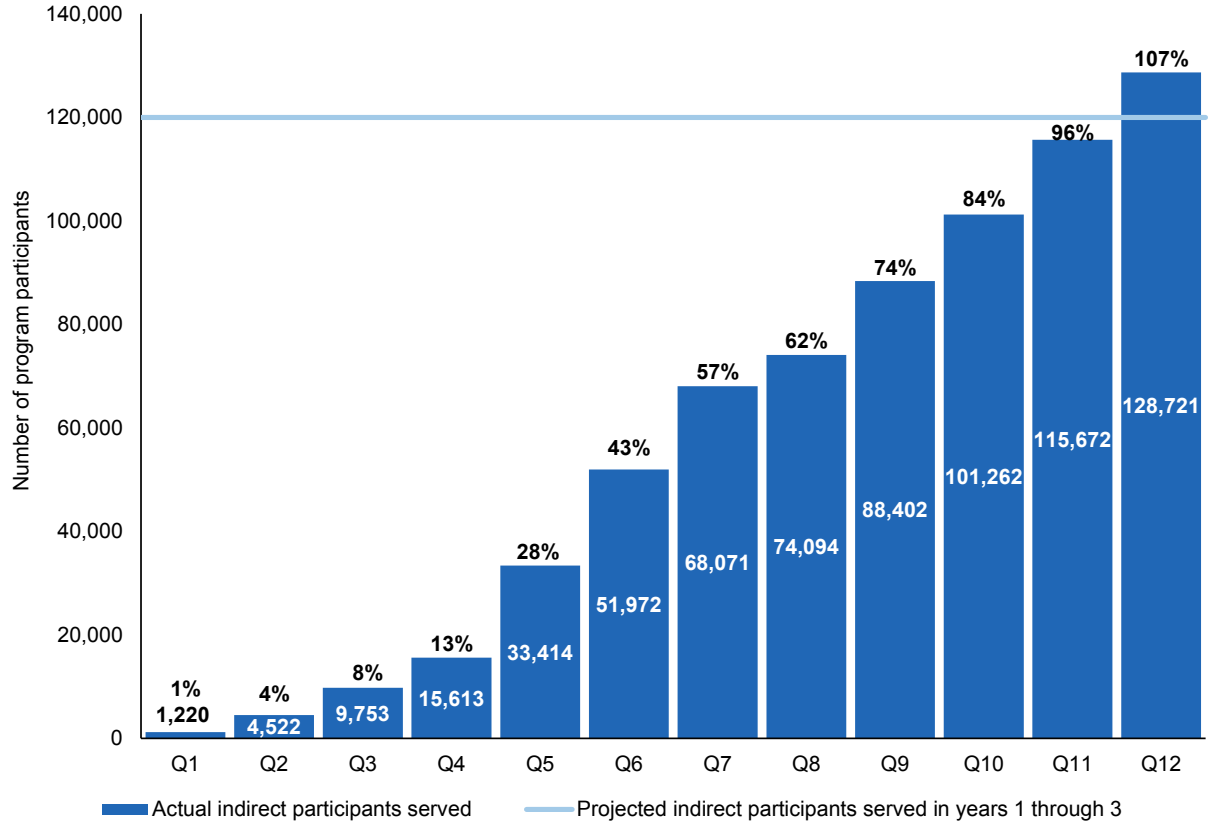
whose PCP submits an eConsult or an eReferral to a participating specialty. Indirect participants are thus passively enrolled into the CORE program. The Association of American Medical Colleges did not experience any major changes to its enrollment process for indirect participants over the course of the cooperative agreement; however, the awardee did lower its enrollment target twice.

The enrollment target was lowered between September and November 2015 from 155,220 to 125,000, and lowered again between September and November 2016 from 125,000 to 120,000. The first change was made after the awardee refined its definition of indirect participants. Initially, the awardee considered the entire primary care population served by a participating AMC as indirect participants, because the entire population was eligible to directly receive an eConsult or eReferral, and even patients who did not get an eConsult or eReferral could benefit from the model indirectly. However, the awardee ultimately defined indirect participants as those patients who received an eConsult as part of their care; this was the reasoning behind the first lowering of the enrollment target. The awardee did not give a reason for the second change to the enrollment target.

b. Evidence of enrollment effectiveness

The Association of American Medical Colleges achieved enrollment effectiveness because it exceeded its three-year enrollment target for indirect participants by the end of the 12th program quarter. The awardee reported that it served 128,721 indirect participants from September 2014 (when it launched its program) through August 2017, which represents 107 percent of its 120,000 three-year projected participants. The awardee lowered its indirect participant enrollment projection from 155,220 to 125,000 in November 2015 and to 120,000 in November 2016.

Figure II.1. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers. The awardee lowered its indirect participant enrollment projection from 155,220 to 125,000 in November 2015 and to 120,000 in November 2016.

c. Barriers and facilitators associated with enrollment effectiveness

Because the progress made by the Association of American Medical Colleges in meeting its three-year enrollment goal for indirect participants was a direct result of the number of eConsults submitted by PCPs, the effectiveness of the enrollment process was primarily influenced by PCPs' willingness to participate in the CORE program and use the eConsult tool.

Although the awardee met its revised indirect enrollment goal, an ongoing barrier during the cooperative agreement was its limited success in encouraging community clinics (as compared to AMC-based clinics) to use the eConsult tool. According to site-level staff, community clinics are typically far away from the AMCs and do not have the same sense of connection to the AMC that the AMC-based clinics do. This limited the uptake of the eConsult by community clinic PCPs. Many of these community clinics also lack wireless Internet connections or have other

technological issues that keep them from using the eConsult tool, which also limited their enrollment of indirect participants.

The participating AMCs implemented different strategies to overcome this challenge. One AMC sought to engage its less responsive community clinics by granting them video access to provider co-management conferences; this was done to enhance their sense of connection to the CORE program. PCP leads at other AMCs made regular visits to their associated community clinics to educate the clinic providers about the value of the CORE program and encourage uptake of the electronic tools. At another site, site-level staff attempted to align their outreach efforts with those of the medical director at the AMC's accountable care organization (ACO) to demonstrate to the community clinics how the CORE program could support the ACO's goals.

2. Delivery of program services

a. Description of and changes to service delivery model

With its HCIA R2 cooperative agreement, the awardee set out to improve and standardize referral and consultation processes between PCPs and specialists. Before the CORE program, referring a patient to a specialist involved either the PCP calling the specialist's practice on behalf of the patient, or the PCP's office giving the patients the contact information to call and make appointments for themselves. This process did not involve in-depth consideration on the part of the PCP about whether the referral was appropriate or necessary. Often, it did not include the transfer of relevant health information from the PCP to the specialist. This sometimes led to superfluous and/or duplicative diagnostic testing. Moreover, this process often omitted any discussion about the co-management roles of the two providers involved in the patient's care or about the duration of specialty care if such care was indeed appropriate. Lastly, the awardee wanted participants to get faster access to specialty care, because normal wait times can range from several days to months. Without prompt access to optimal specialty care, patients can suffer from deteriorating health conditions that may lead to hospitalizations or visits to the emergency department.

In developing the CORE program, the awardee built on the experience and tools that were developed by a similar program conducted by the University of California, San Francisco (UCSF). Under the HCIA R2 cooperative agreement, the awardee implemented an innovative electronic referral (eReferral) and consultation (eConsult) platform for specialty referrals and non-face to-face consultation to improve access to specialty care. The awardee used co-management conferences, departmental meetings, and one-on-one meetings to train PCPs and specialists at the five participating AMCs on the use of the eConsult and eReferral tools.

The awardee hypothesized that using the eConsult tool would help fill gaps in communication and coordination between the primary care and specialty care providers, and facilitate consultation between PCPs and specialists in treating patients. PCPs could choose whether or not to use the eConsult tool when treating participants.

The underlying conditions that warranted use of the eConsult tool were the following: (1) the PCP had a specific clinical question about a patient's condition and/or symptoms that normally would require him or her to refer the patient to a specialist, but might not necessarily warrant an in-person visit, (2) the specialty department that could answer the PCP's question was

participating in the program, (3) the PCP wanted to keep managing the patient's care and only needed some guidance or clarification on how to address a particular issue involved in that care.

The awardee hypothesized that the following outcomes could result from using this tool: (1) fewer unnecessary referrals and visits to specialty providers, (2) increases in patient satisfaction (and better experience with care), (3) reduction of unnecessary or duplicate diagnostic tests and imaging, (4) decreases in patients' out-of-pocket costs and in time spent accessing specialty care, and (5) decrease in ED utilization.

The awardee positioned the eReferral tool to be used at the point of referral, expecting the tool to convey guidance to the PCP on the types of information that specialists needed in advance of the referral and to help the specialists assess whether the referral was appropriate. Information requested in the eReferral template included the basic clinical history for the patient's condition along with confirmation of whether the PCP completed certain key diagnostic and laboratory tests. If an eConsult was converted to an eReferral, the awardee hypothesized that sites could achieve some of the outcomes listed above, including more satisfied patients, faster access to specialty visits, and fewer unnecessary tests and imaging requests. Importantly, the eReferral and eConsult tools can be used independently of each other. The eConsult tool is only converted to an eReferral when a specialist decides to see the patient in person. The eReferral tool replaced sites' previous process for referring patients, and participating providers were expected to use the eReferral tool when they referred patients.

The Association of American Medical Colleges made no major changes to the overall CORE program during the cooperative agreement, nor did it experience any significant delays in its implementation schedule.

b. Evidence of service delivery effectiveness

Overall, the awardee effectively delivered services through the use of the eConsult and eReferral tools and engaged providers to participate in the intervention and use the tools. As described below, providers at participating AMCs gave overwhelmingly positive feedback about their use of the eConsult tool. In addition, the awardee's self-reported monitoring data demonstrated the use of eConsults by PCPs throughout the cooperative agreement. Even though PCP training was not a major component of the CORE program, PCPs' responses to the clinician survey revealed positive feedback about the training they did receive. Although the sites faced issues with the design of the eReferral tool (for example, the length and complexity of templates), they implemented strategies to address these challenges. We provide details below.

Delivery of intervention services. The awardee's self-reported monitoring data and information from interviews suggest that the awardee was effective in delivering program services as intended. Since the program launch, the Association of American Medical Colleges has reported a steady increase in the number of eConsults ordered by PCPs. According to the awardee's self-reported monitoring data, the eConsult rate has increased across participating sites from approximately 4 per 10,000 patients in September 2014 to nearly 14 per 10,000 patients in May 2017. In interviews throughout the course of the cooperative agreement, respondents described the internal quality assurance processes they used to ensure appropriate eConsult submission and response. Some sites also highlighted sample eConsults with appropriate PCP

questions and comprehensive specialist responses during co-management conferences or in the program newsletter.

The eReferral tool met with some resistance from PCPs, however, and it may have been implemented less effectively than the eConsult tool was. In Year 2, for example, interview respondents noted that PCPs thought the eReferral specialty templates were too complicated. Site-level staff reported that PCP concerns may have resulted in less uptake of the tool, which then translated to a decrease in the delivery of services through eReferrals.

Staffing and training. The Association of American Medical Colleges was effective in providing appropriate staffing at the awardee and site level. Respondents brought up few staffing issues during interviews throughout the cooperative agreement, often noting that both awardee and site-level leaders were invaluable to their implementation process. Program staffing at the Association of American Medical Colleges included a project director and assistant project director. In conjunction with its two data partners, Dobson/Davanzo and Viziant, and with UCSF, which piloted an earlier version of the program, the awardee gave participating sites oversight and guidance on the program. Sites used existing staff to oversee site-specific implementation and to deliver program services.

Based on anecdotal evidence shared during interviews, the awardee was effective in training PCPs and specialists on how to use the tools. Given the ease of the technology, and because initiating and responding to an eConsult were relatively straightforward and easy tasks for PCPs and specialists, training them to use the tools was not a major component of the CORE intervention. To train PCPs to use the tools, the awardee first trained PCP leads at each site, who in turn trained PCPs and specialists at their sites during co-management conferences or one-on-one meetings. Interview respondents noted that training went relatively smoothly. Site-level staff gave PCPs and specialists "cheat sheets" to walk them through the EMR steps required for them to use and respond to the tools.

Recruitment and engagement of providers. Overall, the awardee was effective in engaging providers, although participating sites struggled at times to engage busy providers. Although participating sites encouraged providers to use the tools, they sometimes struggled to get busy providers at AMCs and community clinics engaged. Halfway through the third year of the cooperative agreement, the awardee reported that 80 percent of PCPs across all five sites had used the eConsult tool at least once. Although no available measure offers clear insight into how many providers used the tools more than once, interview data suggest that there is broad buy-in from providers and that both eConsults and eReferrals were a normal part of practice for most providers.

> "We have had on the order of 80 percent of PCPs doing eConsults already in the first half or three-quarters of Year 3 at most of our sites. That, I think, speaks to nothing other than the fact that this tool for providing care is something that PCPs find [to be] a useful part of their armamentarium for caring for patients."
>
> —Association of American Medical Colleges interview respondent

Engagement of program participants. The Association of American Medical Colleges did not include direct engagement of program participants as a key feature of the CORE program. Instead, the awardee engaged indirectly with participants through providers' use of the eConsult

and eReferral tools. In fact, a few interview respondents noted that some participants were not even aware their PCP had submitted an eConsult or eReferral on their behalf.

c. Barriers and facilitators associated with service delivery effectiveness

The awardee faced a variety of service delivery challenges in the first two program years (including site leaders with varying levels of commitment to the program and PCPs' frustration with the design of the eReferral templates), but it had largely implemented strategies to overcome these challenges by the third program year.[4] Although some sites continued to face service delivery barriers related to low levels of engagement among community clinics during the third program year, many sites made progress toward increasing providers' use of the eReferral tool.

The facilitators driving service delivery effectiveness were consistent throughout the cooperative agreement. They included: (1) integrating the eConsult and eReferrals tools into existing health IT, (2) leveraging the experience of UCSF during its eReferral and eConsult pilot program, and (3) engaging awardee leaders.

Difficulty with engaging community clinics hindered the effectiveness of the awardee's enrollment processes and service delivery at a few sites. Throughout the cooperative agreement, PCPs at community clinics tended to be less engaged in the program and therefore less likely to use the tools and deliver program services in comparison with providers at other AMC service delivery locations. Although AMC site leads used various strategies to promote engagement among their affiliated community clinics—including co-management conferences, one-on-one conversations between PCP leads and their fellow colleagues, and presentations during grand rounds or faculty meetings—the effectiveness of the awardee's strategies was still limited because distance and technological shortcomings limited the clinics' sense of connection to the awardee. A PCP lead at one AMC site noted that it would have helped if the site could have had (1) a clearer understanding of affiliated community clinics' technological capabilities and (2) strategies for productively involving clinics' local leaders at the onset of program implementation.

In the second program year, PCPs expressed concerns about the design of the eReferral templates. Their chief complaints were that the templates were inefficient, required more information than necessary, and took a long time to complete. As a result, sites focused their attention on streamlining the templates so they were simpler during the third program year. Some sites updated the templates in waves and planned to work through the last wave of updates through the end of the cooperative agreement. At the time of the interviews during the third program year, sites had not yet collected feedback about the updated templates.

In interviews, staff noted that the health IT tools were easy to use, making it easy to deliver services. In particular, integrating the eConsult and eReferral specialty templates into the

[4] Details about the challenges experienced with the eReferral templates in the second year of the cooperative agreement are available in the second annual report, available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

providers' EMR systems facilitated the delivery of CORE services at the point of care and created a direct line of communication between PCPs and specialists.

Sites also said the ability to leverage UCSF's experience piloting the program and adapt UCSF's templates to meet AMC's goals made it easier to deliver services effectively. The specialty templates developed at UCSF, for example, were the basis for those used in the participating sites. Site-level leaders also reported that specialty templates were delivered to site-level staff in a way that made them easily adaptable to site-specific workflows. The awardee also leveraged UCSF's experience by including UCSF staff in awardee-sponsored office hours to answer sites' EMR-related questions.

Finally, program and site-level staff said that the engagement of senior leaders at both the Association of American Medical Colleges and at participating sites was a critical facilitator of service delivery. For example, senior leaders at the Association of American Medical Colleges worked directly with leaders at one participating site to help them identify strategies to overcome the limited engagement of providers at affiliated community clinics. Senior leaders at the Association of American Medical Colleges also worked directly with leaders at another participating site to help them manage disruption and maintain CORE program momentum during a site-level leadership transition. Interview respondents also noted that senior leaders at sites, such as executive sponsors and senior specialists, "talked up" the program to junior faculty and used the technology themselves, which encouraged engagement among PCPs and specialists.

C. Assessment of perceived program effects on the delivery of care and outcomes

Clinician survey and interview respondents alike overwhelmingly believed the CORE program had positively affected the delivery of care, specifically noting that the program had improved collaboration and communication between PCPs and specialists and thereby enhanced the quality of care they provided to participants. Highlights of the positive perceptions of the program's impacts on service delivery included:

- Ninety-one percent of surveyed PCPs believed the CORE program had led to better patient care.
- Eighty-eight percent of survey respondents reported that the program had a positive impact on the quality of care and services they provided.
- Eighty-eight percent of survey respondents said the program increased collaboration between PCPs and specialists.
- Eighty-two percent of survey respondents said they believed CORE had a positive impact on care coordination.
- Eighty-six percent of survey respondents reported that the program increased PCPs' knowledge about issues that often require specialist input and/or referral

Feedback from interview respondents at different sites corroborated these findings, and included examples of PCPs applying what they had learned from specialists' input on eConsults

to similar cases that followed. Interview respondents also said PCPs were more comfortable reaching out to specialists as a result of the CORE program.

Respondents to the clinician surveys and interviews also believed the CORE program had positively affected outcomes including indirect participants' (patients) satisfaction with care, access to specialty care, and ED utilization. Highlights of the responses on care delivery included:

- Ninety percent of clinicians who completed the survey believed the program increased participants' access to specialists.
- Eighty-seven percent of respondents reported that the program had a positive impact on their ability to respond promptly to patients' needs.
- Seventy-six percent reported that the program had a positive impact on participants' satisfaction with care.

Site-level staff corroborated these findings in interviews and noted that participants were pleased about not waiting a long time for specialist input and avoiding the cost of a specialist visit.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Although leaders at the Association of American Medical Colleges, along with staff at participating sites, thought the CORE program was positively influencing the delivery of care and participants' outcomes, some implementation issues do have implications for the interpretation and analysis of program impacts.

Given the nature of the CORE program, participants may benefit from exposure to just one eConsult during one office visit. In fact, participants may benefit from the program without exposure to any eConsult at all if their PCP applies the knowledge gained from a specialty provider during a previous eConsult to the care he or she gives to other participants.

In addition, a possible barrier to the interpretation of an impact analysis of the CORE program is that it will be difficult to attribute any outcomes that are achieved, such as higher levels of participant satisfaction or decreased use of the ED, to the program because of other confounding factors that may impact these outcomes and others.

Similarly, if an impact analysis reveals that the program was associated with lower overall health care costs at participating sites, it will be difficult to interpret those results or attribute them directly to the program. One interview respondent, for example, said it was possible that increased access to specialty care because of the CORE program may not translate into overall decreases in service utilization or costs, because preventing unnecessary specialist visits for some participants opens up providers' availability to serve other, potentially more complex patients—whose care could drive up site-level service utilization and costs.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of the Association of American Medical Colleges' CORE program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Association of American Medical Colleges

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	100,643[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2016	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	5,544
Likelihood of all-cause hospitalizations	2,758
MDE sample size requirement to detect 20% effect	
Total expenditures	1,386
Likelihood of all-cause hospitalizations	690
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	Some issues, but these appear to be surmountable; expect to select a comparison group
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician survey
Implementation data that will be analyzed	None

[a] The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We plan to carry out a rigorous impact analysis of the e-Referral and e-Consult intervention. As a first step in this analysis, we will construct a comparison group of patients from a matched set of physician practices associated with academic medical centers. To identify candidate practices for matching, we have secured information on the characteristics of academic medical centers from IMS Health. The characteristics include specialty, number of providers, payer mix, and number of patients. We are currently matching physicians to these practices by using the CMS National Plan and Provider Enumerations System (NPPES). From this set of potential matches, we will construct a final sample of practices matched on payer mix, total number of providers, number of PCPs, ratio of physician to nonphysician providers, and number of Medicare patients served by PCPs.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents a summary of the baseline characteristics of the treatment group, measured during the 12 months before the enrollment date for each beneficiary who visited a primary care clinic within a participating AMC network. For the purpose of our evaluation, the treatment group consists of individuals who visited primary care clinics that were exposed to the eCR program and were part of a participating AMC network: (1) Dartmouth-Hitchcock; (2) the University of California, San Diego; (3) the University of Iowa; (4) the University of Virginia; and (5) the University of Wisconsin.

The five participating AMC networks launched their eCR programs in September 2014. By the end of May 2016, 102,558 Medicare beneficiaries had visited one of the 1,320 primary care providers identified by the awardee as participants in the program. [5] Subsequently, 11,494 beneficiaries were excluded from the analysis because they were not enrolled in Medicare FFS.

In presenting baseline characteristics for this report, we restricted the treatment group to Medicare beneficiaries who (1) visited primary care physicians associated with the five AMCs after September 1, 2014; (2) were enrolled in Medicare FFS, Parts A and B, with Medicare as the primary payer at the time of their enrollment date with one of the five AMCs; (3) had met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment); and (4) who could be identified in the Virtual Resource Data Center's Medicare health insurance claim-to-beneficiary ID crosswalk. The calendar period covered by the baseline quarters is determined by the enrollment date for each participant and therefore varies by participant. After excluding 4,078 beneficiaries who were not enrolled in Medicare FFS and 7,416 who were not enrolled for at least 90 days prior to the program enrollment date,

[5] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

we included a total of 91,064 participants in the analysis of baseline characteristics for this report.

Our analysis of baseline demographic characteristics (Table III.2) indicates that compared to the general Medicare population, the Medicare FFS beneficiaries who saw a participating primary care physician have similar age and sex composition but include smaller percentages of Hispanics and African Americans.[6] In terms of reasons for Medicare eligibility, the proportion of patients in the treatment group with disability insurance benefits is higher than the one for all Medicare beneficiaries, while the hierarchical condition categories (HCC) score of program participants is also higher than the national average, with a mean of 1.28.

Table III.3 shows baseline expenditure and utilization data for a common set of measures for the four quarters prior to enrollment. The higher-than-average HCC scores are consistent with the high total Medicare expenditures per beneficiary per month (PBPM) shown in the table. On average, the PBPM expenditures for the treatment group during the 12 months before enrollment were $1,109—higher than the national average Medicare expenditure of about $790 per month. Multiple spending categories drove the high costs, including inpatient ($472 PBPM), outpatient ($247 PBPM), and physician services ($206 PBPM).

Fourth-quarter expenditures, spanning the three months before enrollment, were higher than earlier quarters ($1,621 PBPM in the fourth quarter). In this quarter closest to enrollment, average expenditures were highest across most expenditure categories—and notably so for acute inpatient ($848 PBPM and 80 percent higher than the 12-month average), physician services ($246 PBPM and 19 percent higher than the 12-month average), outpatient ($271 PBPM and 10 percent higher than the 12-month average), and skilled nursing facility ($102 PBPM and 42 percent higher than the 12-month average).

Increases in measures of utilization accompanied the rise in fourth-quarter expenditures, with the exception of primary care visits. The average rates per 1,000 beneficiaries over all 12 baseline months included 358 acute hospital admissions; 610 outpatient ED visits; 65 observation stays; and 8,536 specialist visits. The rates of utilization were highest for the enrolled population in the quarter closest to enrollment across all the utilization types listed above. Rates of acute hospitalization, outpatient ED visits, and observation stays in the fourth quarter were 54 percent, 24 percent, and 29 percent higher, respectively, than the 12-month averages. Similarly, the percentages of beneficiaries with a hospital admission (10 percent), an outpatient ED visit (13 percent), or an observation stay (2 percent) are the highest in the quarter closest to enrollment, in comparison to the three previous quarters. The increases in these utilization measures suggest that the enrolled patients were incurring hospital-related charges just prior to the enrollment event.

[6] See http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf?sfvrsn=0.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the Association of American Medical Colleges' program through May 31, 2016

	All participants (N = 91,064)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	16,927	19
65 to 74	41,745	46
75 to 84	22,315	25
85 and older	10,077	11
Gender		
Female	51,993	57
Male	39,071	43
Race		
White	80,579	88
Black	5,503	6
American Indian, Alaska Native, Asian/Pacific Island American, or other	2,788	3
Hispanic	803	0.88
Original reason for Medicare eligibility		
Old age and survivor's insurance	67,509	74
Disability insurance benefits	22,332	25
End-stage renal disease (ESRD)[a]	1,223	1
Hospice[b]	291	0.32
Medicare/Medicaid dual status, percent dual[b]	15,844	17
HCC score[c]		Statistic
Mean		1.28
25th percentile		0.47
Median		0.81
75th percentile		1.48

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Note: The baseline year is defined as the 12 months before each beneficiary's enrollment date. The enrollment date is defined as the date on which we first have evidence that a beneficiary received a service at a participating facility between 9/1/2014 to 5/31/2016. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the Association of American Medical Colleges' program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	91,064	84,897	87,444	90,792	91,059
Average Medicare expenditures PBPM[a]					
Total	1,109 (10)	888 (14)	907 (12)	986 (32)	1,621 (20)
Acute inpatient	472 (7)	319 (11)	333 (8)	364 (31)	848 (15)
Inpatient other[b]	46 (2)	29 (3)	34 (2)	41 (3)	77 (4)
Outpatient[c]	247 (3)	232 (3)	233 (3)	249 (3)	271 (3)
Physician services	206 (2)	192 (3)	184 (2)	200 (2)	246 (3)
Home health	31 (<0.5)	27 (1)	29 (1)	30 (1)	37 (1)
Skilled nursing facility	72 (1)	56 (2)	61 (2)	68 (2)	102 (3)
Hospice	10 (1)	7 (1)	8 (1)	10 (1)	13 (1)
Durable medical equipment	25 (1)	25 (1)	25 (1)	25 (1)	26 (1)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	358 (3)	278 (7)	285 (7)	302 (17)	552 (6)
Outpatient ED visits	610 (7)	537 (9)	549 (9)	586 (10)	756 (10)
Observation stays	65 (1)	58 (2)	55 (2)	60 (2)	84 (2)
Primary care visits in any setting	4,864 (21)	4,961 (28)	4,836 (28)	4,904 (38)	4,745 (32)
Primary care visits in ambulatory settings	4,049 (15)	4,345 (21)	4,176 (21)	4,183 (24)	3,509 (19)
Specialist visits in any setting	8,536 (43)	7,894 (51)	7,878 (51)	8,318 (91)	9,933 (72)
Specialist visits in ambulatory settings	6,293 (27)	6,249 (34)	6,178 (34)	6,403 (50)	6,311 (32)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	21 (< 0.5)	6 (< 0.5)	6 (< 0.5)	6 (< 0.5)	10 (< 0.5)
Percentage with an outpatient ED visit[e]	29 (< 0.5)	9 (< 0.5)	10 (< 0.5)	10 (< 0.5)	13 (< 0.5)
Percentage with an observation stay[f]	6 (< 0.5)	1 (< 0.5)	1 (< 0.5)	1 (< 0.5)	2 (< 0.5)
Percentage with a 30-day readmission among all discharges	19 (< 0.5)	18 (1)	18 (1)	18 (< 0.5)	20 (< 0.5)
Percentage of participants with a readmission among all participants	3 (< 0.5)	1 (< 0.5)	1 (< 0.5)	1 (< 0.5)	1 (< 0.5)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Notes: The baseline year is the 365 days before each participant's enrollment date. The participant's enrollment date is defined as the date of the first Medicare-covered visit after the start date of the physician's participation in the project. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Through its HCIA R2 cooperative agreement, the Association of American Medical Colleges did not develop a new payment model, but instead pursued options under existing FFS payment. The awardee provided a new service through the eConsult tool and identified different ways to reimburse for this service, including developing new billing codes and including the eConsults as a qualifying service within an APM.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

The awardee targeted both public and commercial payers to reimburse for use of the eConsult tool. The awardee proposed a new payment approach in which the specialists responding to eConsults would bill by using a "G code." A G code is a temporary code that CMS approves and uses to pay for certain services; it can be activated for restricted populations (for example, only providers at the participating AMCs).

Once the awardee receives approval for the G code, it can immediately obtain reimbursement for its services because the G code circumvents the need to identify current procedural terminology (CPT) codes. The awardee stressed that the latter is a difficult and time-consuming process. In addition, because G codes often can become permanent codes, there is a path open for sustaining this payment approach. Once an AMC receives the specialist payment from CMS, it would then split the payment and reimburse both the specialist and PCP.

The awardee also hoped to persuade a private payer to pilot a reimbursement system for the eConsults and to assess the impact of the tool. This approach was influenced by UCSF's experience: UCSF contracted with commercial payers and the self-funded plan covering its employee population to reimburse the specialists for using the eConsult tool.

The awardee's partner, Dobson/Davanzo, built a return-on-investment (ROI) model to help drive discussions with private payers about reimbursing providers for the time dedicated to replying to eConsults. In conjunction with Dobson/Davanzo, the awardee assessed potential cost savings to payers for reimbursing eConsults by analyzing program data on the decrease in referral rates to a specialist by a PCP. The awardee noted that there would need to be parameters in place to ensure no overuse of eConsults. There is no set fee amount that's been determined for private payers, because each AMC will negotiate that amount with the payers. Patients would continue to pay the regular co-pay required by their insurance plans for a PCP visit.

The awardee is also looking into an alternative payment approach that aligns the eConsult service with current CMS payment approaches, such as APMs. The awardee is looking into

whether PCPs requesting the eConsults could bill with prolonged service CPT codes under a CMS waiver if the PCPs participate in a larger APM, such as an accountable care organization. This option would create a second version of the proposed payment approach, in which a specialist within the AMC would be reimbursed using the G code, and the PCP in a community-based practice outside of the AMC would be reimbursed under the APM waiver. This version would allow PCPs who are not within the AMCs' system to use the service delivery model.

C. Status of the payment model

The Association of American Medical Colleges plans to use the G codes once they are approved by CMS. The awardee is currently working on refining its payment approach and discussing internally whether the alternative option is feasible. At the same time, the awardee is engaging key stakeholders to help identify the best path for implementation, including meeting with the American Medical Association's Specialty Society Relative Value Scale Update Committee (RUC) to discuss creating new CPT codes to replace the temporary G code if needed. The RUC is a multispecialty physician committee that provides recommendations to CMS on the resources and codes that are needed to provide medical services and that are integral to updating CPT codes.

None of the sites have executed a contract with commercial payers, but three of the AMCs now have arrangements for reimbursement of eConsults for their self-insured employee population. All sites are exploring opportunities to expand this to other payers during the no-cost extension period. In the meantime, the AMCs are providing the participating primary care provider and specialist eConsultant a 0.5 relative value unit (RVU)[7] or equivalent (based on local compensation models) for a completed eConsult. This is a continuation of the practice in place during the cooperative agreement—because the partner AMCs had physician compensation plans in place that reward productivity, they provided RVU credits and monetary reimbursement to incentivize participation in the CORE program. This payment recognizes the time it takes for the specialist to review and respond to the eConsult, and for the PCP to implement the specialist's recommendations in lieu of a referral. These payments will continue to be self-funded by the AMCs.

D. Factors associated with the development of the payment model

In awardee documents and interviews, the awardee noted that its partnerships with two key stakeholders, UCSF and Dobson/Davanzo, have helped it find ways to reimburse for eConsults. The awardee continues to leverage UCSF's data and experiences to identify opportunities to work with CMS and regional payers to recognize eConsults as a reimbursable service. The Association of American Medical Colleges' partnership with Dobson/Davanzo, which had extensive experience in payment model development as noted by the awardee, helped the awardee identify a way to reimburse for eConsult and identify an option for aligning the reimbursement with CMS' payment approaches.

[7] RVUs are a measure of physician work often used in systems for paying physicians. See "How Medicare Pays Physicians," *The New York Times*, December 3, 2010. Available at http://economix.blogs.nytimes.com/2010/12/03/how-medicare-pays-physicians/?_r=0.

The awardee noted that one hindrance to developing the payment approach was that the FFS environment lacks a precedent for reimbursing both PCPs and specialists for their efforts in coordinating services and executing a care plan. Through discussions with its partners, CMMI, and HCIA R2 consultants, the awardee realized that the focus of reimbursement would rely on specialists billing for services and the payment could be shared between the specialists and the PCP. Another challenge stemming from lack of a precedent is setting up payment for providers who are not employed by the AMC. This is one reason the awardee is looking into payment options that allow external providers to participate in an APM and still have access to consultation from the AMC's specialists.

The final challenge mentioned by the awardee is the feedback from CMS expressing no interest in creating new FFS codes. The awardee felt that this limited their ability to identify which codes could be used for specialist reimbursement. They tried to use existing CPT codes but realized that the codes they wanted to use required both verbal and written communication between specialist and PCPs. Since the eConsult model is based on electronic communication, those codes were not a good fit.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

As they awaited further progress and expansion of the payment options to secure long-term sustainability of the CORE program, the awardee and its implementing sites believed they would be able to sustain the program in the short term due to sufficient internal support for the program. In addition, the awardee's 12-month no-cost extension would give the Association of American Medical Colleges more time to prepare and gain support for the payment model. During the three-year cooperative agreement, the awardee scaled the program to other member AMCs, affiliated providers, and patient populations. It had plans to continue expanding the program beyond the cooperative agreement. The awardee did not report plans to replicate the program to external, non-AMC sites or providers.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

- This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:
- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

During the second program year, the Association of American Medical Colleges was actively working to develop its sustainability plan and had begun implementing some of its strategies. The awardee started analyzing cost savings, patient satisfaction, and clinical outcomes to help build support from leaders and payers for sustaining the model. The awardee also conducted in-person interviews with leaders at each implementing site, which helped gain buy-in

from the site leaders and got them to identify next steps for activities designed to sustain the program at their site.

In addition, the awardee worked with the American Medical Association to identify different ways to reimburse providers for use of the eConsult tool; started training new physicians to view the program as embedded in their regular work flow; and facilitated a collaborative for implementing sites to share approaches for sustaining the program with each other. Further, many of the implementing sites were planning to scale the program to more medical specialties. The Association of American Medical Colleges was scaling its program to seven other AMC members in the second program year and had plans to continue expanding the program at existing implementing sites, as well as at new AMC member sites. The awardee did not report plans to replicate the program to external, non-AMC sites or providers.

C. Implementing the SSR plan: progress and changes

Sustainability. Continuing to implement and expand the payment options (described in Chapter IV on the payment model) was the awardee's key strategy in Year 3 to sustaining the program in the long term. As part of its 12-month no-cost extension, the awardee planned to focus on (1) gathering and analyzing data needed to demonstrate the program's positive effects on patient care and costs, and (2) implementing the payment options.

Even without a formal analysis that showed a positive ROI, the HCIA R2 cooperative agreement enabled implementing sites to demonstrate the program's value to providers and patients alike. As a result, all five participating sites reported plans to use internal resources to sustain program components to varied degrees and through varied means. For example, one implementing site expected its board to approve integration of the eConsult and eReferral tools into its broader telemedicine department, which would allow it to more directly access tele-health resources including health IT and staffing. The site also planned to retain two positions to lead the program. Another site planned to incorporate the program functions into its regular day-to-day operations by embedding the tool into its EMR and transitioning specific functions to quality management teams focused on billing and prescription. The primary care lead would transition to the IT department in an "informaticist" position and take responsibility for the program's overall monitoring and reporting activities. The site also plans to continue making incentive payments to specialists.

Scalability. The five implementing sites were working to scale the program in different ways, including adding specialty-to-specialty eConsults, as well as expanding the program to pediatric specialties and external primary care sites from which the AMCs receive referrals. Specific scaling strategies are described in more detail below:

- One AMC expanded the program to 30 specialties, more than any other site. The same site was in the process of expanding the program to its children's hospital.
- One AMC was building teams to expand program services to pediatric patients. This was primarily driven by a new state mandate for large AMCs to make their pediatric specialty expertise and services more available to other providers throughout the state. The AMC also was working with a few small community hospitals that could benefit from eConsults, in part due to their distance from the AMC, which would charge a fee for the service.

- One AMC was more focused on scaling the program to the external primary care groups that it receives specialty care referrals from and that currently do not use the same EMR (Epic). The AMC also expected the need for specialty care to increase as it expanded its own primary care clinics; expanding the program internally would avoid the need to build new specialty clinics to meet that demand.

Further, the Association of American Medical Colleges was scaling the program by implementing it at more member AMCs. It launched a second cohort of seven AMCs in the spring of 2016. Each of those sites had "gone live" with between 2 and 12 specialties, with further expansion of specialties ongoing. The awardee also was recruiting a third cohort of between three and six additional AMCs. Three AMCs had already committed and several had expressed interest, but the awardee did not want to overextend itself by adding more than six new sites. Because the Association of American Medical Colleges would not have financial resources through HCIA R2 after the cooperative agreement ended, the new AMCs were self-funding the program's implementation and therefore had more latitude in how they implemented the program. However, the Association of American Medical Colleges made recommendations—for example, on the minimum number of staff needed to support the program—and believed that the sites were implementing the program with fidelity to the original model.

Based on its experience with the second cohort, the awardee found that initial program implementation can take place quickly, but sites often need or want extra time to stand on their own; therefore, the third cohort would receive two-year (instead of one-year) contracts. The sites were trying to determine how much review and oversight remain necessary as the program matures, and whether it is important to automate or streamline the oversight of the program's quality.

Replicability. Neither the awardee nor the sites reported plans to replicate the program to non-AMC sites or providers.

D. Factors associated with progress toward implementing the SSR plan

The CORE program was embedded into each implementing site's EMR system and had strong support from the awardee and the five sites. Respondents expected this to foster CORE's sustainability. One respondent explained that, even without robust ROI, leaders believed that addressing medical issues through an eConsult between PCPs and specialists is more efficient than an in-person visit or a phone call. This assumption of cost savings was key to maintaining the program while the awardee worked to institute an ongoing payment approach. As the awardee lead summarized, "All five of our sites, despite not having payers on board for the most part, are fully committed to sustaining the model because the model provides better care, is more efficient for their providers even if it doesn't bring in more revenue, it's good for patients, the providers like it, and there would be some revolt if it went away at this point. Most of our sites are continuing to want to grow." Further, the ongoing growth of the program and interest from other AMCs in adopting it gave the awardee an argument to make to its leaders for maintaining the internal staff to support the program and help additional sites implement it.

However, longer-term sustainability and significant scaling and replication were more uncertain and faced three main challenges: (1) demonstrating a positive ROI; (2) receiving payment from commercial payers; and (3) ensuring ongoing quality. As noted before, the awardee was unable to gather and analyze data necessary to demonstrate a positive ROI before the end of the cooperative agreement, although it hoped to be able to do so during the no-cost extension. Program leaders needed to continually demonstrate that the program was good for patients and reduced overall costs for the system, but not all implementing sites had sufficient data to compare the cost of an eConsult to that of an in-person visit. Scaling the program to external providers was seen as a way to balance this tension, because it should result in more referrals to the site that might have gone elsewhere in the past. And the potential movement to value-based payments would offer more flexibility to deliver services, possibly making a discrete payment for the program less important over time.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in the Association of American Medical Colleges' CORE program. Medicare beneficiaries who received primary care services from primary care practices affiliated with non-participating AMCs, similar to the five awardee AMCs, will comprise the comparison population. Each of the five awardees will be individually matched to a comparison AMC. Characteristics considered when matching AMCs include patient volume, geographical location, ACO membership, relative share of primary care providers and specialists in areas associated with the intervention, and mix of beneficiaries served by the primary care providers in those health systems. We will develop a matched comparison group and provide informal memoranda to CMS showing matching results in the form of a balance plot and intermediate results from impact estimates between January 2018 and June 2019.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: American College of Cardiology Foundation

April 18, 2018

Dana Jean-Baptiste (Mathematica Policy Research)
Shannon Heitkamp (Mathematica Policy Research)
Thomas Grannemann (Mathematica Policy Research)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries. The American College of Cardiology Foundation received a six-month, no-cost extension to complete claim and clinical data analysis and to collect the final three months of follow-up survey data.

The federal evaluation has two interrelated goals. The first goal is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

The American College of Cardiology used funding from HCIA R2 to create the SMARTCare program, which it implemented at nine clinical sites throughout Wisconsin and Florida. The program relied on health information technology (health IT) to (1) provide clinical decision support for managing stable ischemic heart disease (SIHD) to cardiologists and other clinical specialists at the point of care, (2) support patient-clinician shared decision making, and (3) enable the use of clinical registries to track and improve care. The key innovations within SMARTCare's design were (1) the bundling of five related decision-support tools that use evidence-based medicine and (2) a bundled approach to paying for services (see Table I.1). The awardee intended that the tools would guide clinician decisions—from ordering tests to performing procedures—in order to reduce inappropriate use of cardiac screening tests and procedures and to reduce rates of complications. The tools also provided customized, patient-specific estimates of the risks and benefits of specific procedures, as well as educational materials to support shared decision making.

Table I.1. SMART Care Tools

Tool	Functionality in the SMARTCare program
IndiGO	Calculates and displays an individual patient's risk of an adverse event and suggests and prioritizes approaches with the greatest potential to reduce that risk through an interactive display used with the patient.
FOCUS	Computerized decision-support tool that incorporates patient-specific information to determine whether ordered imaging meets appropriate use criteria established by the American College of Cardiology.
ePRISM and eLumen	ePRISM produces a customized, patient-specific consent form that provides participant education along with estimates of the benefits and risks of complications, tailored to each participant prior to invasive cardiac catheterization or percutaneous coronary intervention. eLumen is an extension of ePRISM that uses the patient's information to guide the clinician during catheterization to reduce complications and potentially improve resource use.
Tonic	Web-based application that runs on an iPad and is used to document patient consent for collection of personal health data using either the Seattle Angina Questionnaire-7 or Heart Quality of Life surveys and to track patient-reported outcomes going into and coming out of treatment.
Health Dialog	Provides additional patient education materials—such as a pamphlet explaining stress chest discomfort—to inform and prepare patients for a more effective dialogue with their clinicians.

Table I.2. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The American College of Cardiology Foundation (ACCF) focused on changing clinician behavior by providing (1) decision-support tools at the point of care to assess treatment options for SIHD, (2) patient education materials on specific treatment options, and (3) individually tailored risk and benefit information to support shared decision making.
Major innovation	ACCF's major innovations were the use of health IT to bundle evidence-based, decision-support tools as well as the awardee's bundled approach to paying for services.
Program components	• Clinical decision support • Shared decision making • Health IT
Target population	Participants in the SMARTCare program were Medicare beneficiaries whose clinicians provided care and evaluated patients for SIHD by using the SMARTCare tools at nine clinical sites. For that reason, participants are indirect participants.
Theory of change/ theory of action	Improving risk communication and shared decision making between participants and cardiac clinicians would lead to optimizing medication and lifestyle programs for the greatest potential impact on a participant's risk factors.
Payment model	Value-based payments, bundled or episode payment
Award amount	$15,830,092
Effective launch date	11/11/2014
Program setting	Provider-based settings: Primary care clinic, specialty care clinic, hospital, academic medical college setting
Market area	Rural, urban, suburban
Market location	Florida, Wisconsin
Target outcomes	• Decrease in the percentage of imaging tests that do not meet appropriate use criteria • Decrease in the percentage of elective percutaneous coronary interventions (PCIs) that do not meet appropriate use criteria • Decrease in the risk-adjusted bleeding complication rate for elective PCIs • Improvement in either the Seattle Angina Questionnaire score (patients with chest pain) or the Heart Quality of Life score (patients without chest pain) • Increase in adherence to coronary artery disease (CAD) treatment guidelines: – Angiotensin converting enzyme inhibitor or angiotensin receptor blockers therapy prescribed for participants with diabetes or left ventricular systolic dysfunction (LVSD) – Oral antiplatelet therapy prescribed for participants with CAD – Aspirin or other antithrombotic prescribed for participants with acute myocardial infarction, coronary artery bypass graft, PCI, or LVSD – Lipid control prescribed for participants with CAD – Beta-blocker therapy prescribed for participants with CAD – Tobacco use assessment and tobacco cessation counseling administered to participants

[a]After the initial planning period, the awardee's program became operational as of this date.

IT = information technology; SIHD = stable ischemic heart disease.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was partly successful in implementing its program by the end of the initial three-year cooperative agreement period. We based this conclusion on several factors. First, the awardee enrolled 29,053 participants—108 percent of its enrollment target—by the end of the initial cooperative agreement. However, a small number of clinicians seem to be driving enrollment across all sites, suggesting a lack of clinician buy-in. The lack of clinician buy-in stemmed from two sources: (1) difficulty using the tools due to interoperability issues within sites' electronic medical record (EMR) systems and (2) concerns that the tools would not have an impact on patient care. Participating clinicians and other implementation staff reported that the program had a positive effect on care delivery but that there were many operational issues that prevented the program from being fully successful—including, adverse payment incentives for clinicians operating in a fee-for-service (FFS) environment.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of the American College of Cardiology's SMARTCare program, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. The American College of Cardiology designed a condition-specific, bundled episode payment to support diagnosis and treatment of patients with SIHD. The payment model is still in development.

Sustainability plans. The American College of Cardiology reported that almost all implementing sites would sustain some aspects of the SMARTCare program. The awardee did not report plans to scale or replicate the program.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey, which was fielded from July 2016 to October 2016, achieved a response rate of 83 percent with a sample size of 31 respondents. The clinician survey, which was fielded from March 2017 to June of 2017, achieved a response rate of 57 percent with a sample size of 263 respondents. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partially successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals. The American College of Cardiology made no significant changes to the overall SMARTCare program nor did it experience any significant delays in its implementation schedule.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

SMARTCare participants are patients whose clinicians are using the SMARTCare tools, as detailed in Table I.1. Once the clinician uses the FOCUS tool to test for appropriateness of an imaging test, the patient is automatically enrolled into the program. Patients must consent to have their demographic characteristics and medical information included in the data analyses that will evaluate the program, but patients may receive some program services without consenting to share their data.

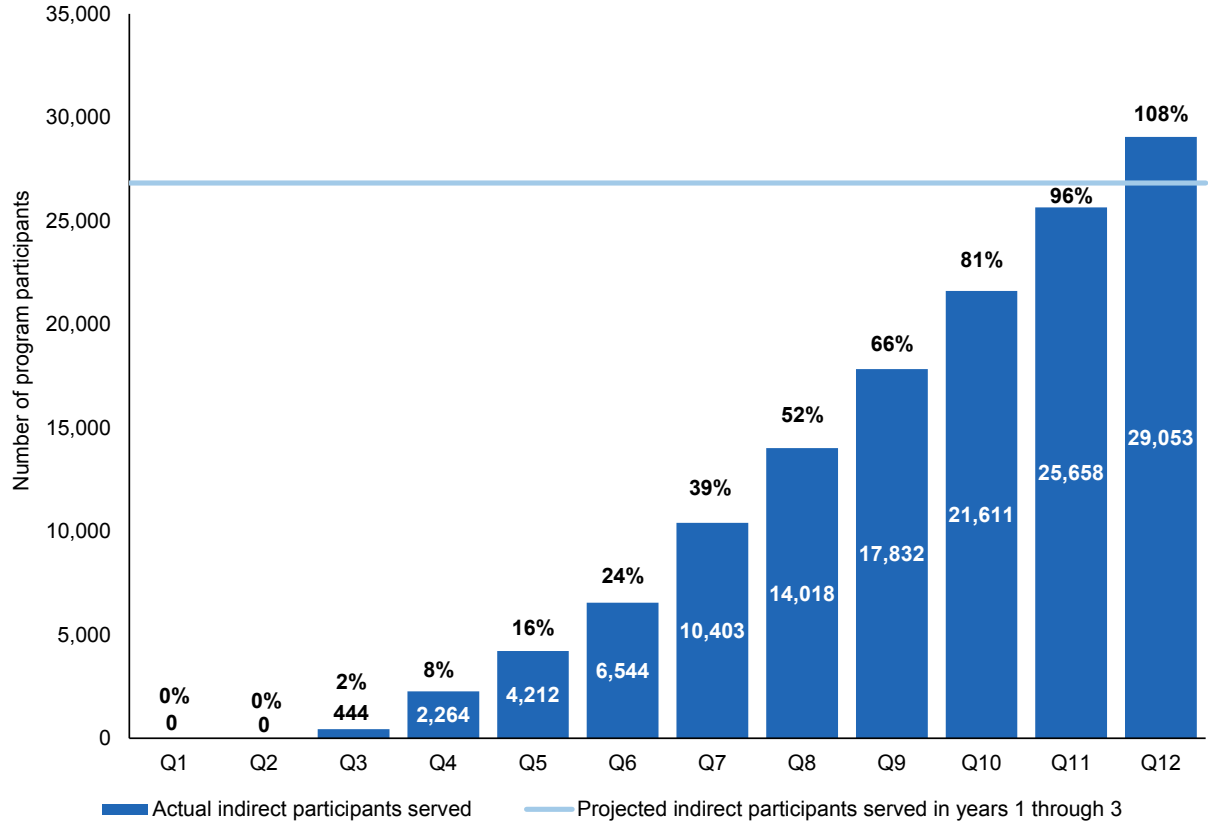
The American College of Cardiology did not experience any major changes to its indirect participant enrollment process during the cooperative agreement. However, the awardee did

change its enrollee target twice. The first time was during the second year of the cooperative agreement because it misclassified program participants as direct participants. The American College of Cardiology lowered its target from 84,583 participants to 25,557 participants once it started correctly reporting indirect participants. The awardee then increased its indirect participant target to 26,832 in the third year because it saw increased use of the FOCUS tool. This increased use likely reflected the awardee's efforts that year to engage clinicians in SMARTCare in order to meet its indirect participant enrollment goals.

b. Evidence of enrollment effectiveness

The American College of Cardiology was effective in meeting its enrollment target by the end of the last program quarter. Overall, the awardee reported that it indirectly enrolled 29,053 participants from March 2015 through August 2017—about 108 percent of its revised 26,832 three-year projected participants. The awardee lowered its target from 84,583 participants to 25,557 participants in Year 2, then increased it to 26,832 participants in Year 3. After the first year of the cooperative agreement, the awardee enrolled 2,264 indirect participants. During the second year of the cooperative agreement, the awardee hit its stride and cumulatively enrolled 14,018 indirect participants. Although exposure to the SMARTCare program can be relatively brief (the exposure period ranges from being evaluated for SIHD to potentially receiving procedures related to the patient's condition), there was significant variation in the number of participants enrolled across clinicians and sites. Feedback from interview respondents indicated that this variation may be partly because of provider engagement, but it also may be due to the fact that a cardiology clinic would be expected to enroll more patients than a site with a different provider mix.

Figure II.1. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers. The American College of Cardiology lowered its target from 84,583 participants to 25,557 participants in Year 2, then increased it to 26,832 participants in Year 3.

c. Barriers and facilitators associated with enrollment effectiveness

The American College of Cardiology's progress in meeting its three-year enrollment goal was influenced primarily by patient consent and clinician engagement. Although the awardee was able to meet its enrollment goal by the end of the third year of the cooperative agreement, it sometimes struggled to address these challenges.

The documentation of patient consent continued to be a barrier to both enrollment for patients and participation by clinicians. The consent forms were seen as burdensome for one site because the clinician was required to document patient consent by using an internal site-specific form and a SMARTCare form, thus duplicating the time and effort required to enroll patients. This duplicative effort turned some clinicians away from the program because they were too busy to handle additional paperwork. Another concern mentioned in the first two years was

patient mistrust of data collection efforts. Site-level staff made an effort to discuss with patients what type of data would be collected and how it would be used.

Issues with clinician engagement and buy-in recurred throughout the cooperative agreement. Clinician buy-in was critical to the program because patients were indirectly enrolled after a clinician used the FOCUS tool. Site-level staff emphasized that engaging clinicians through various efforts, including discussions with leaders about the tools and presentations at departmental meetings, would lead to buy-in and use of the FOCUS tool for patients. However, low clinician engagement was a significant factor that impacted enrollment. In Year 1, site leaders demonstrated varying levels of support and encouragement for the program. Engaged site leaders motivated staff to use the tool for their patients and ensured that resources to support implementation were available. Comparatively, disengaged leaders were not convinced SMARTCare would meet future demand from payers or drive decision support. In Year 2, some clinicians continued to be unconvinced that the decision-support tools were necessary. They remained resistant to using the tools because they did not think the data derived from the tools were sufficiently reliable to drive quality improvement efforts. In Year 3, although American College of Cardiology leaders noted that clinicians tended to be more engaged when they saw data from the tools, they were not able to share financial data with clinicians because of issues with obtaining Medicare data from the Research Data Assistance Center (ResDAC) and all-payer data in Wisconsin from the All-Claims Payer Database (Florida does not have a similar database). Awardee leaders hypothesized that data from ResDAC would have addressed site administrators' concerns about whether the program could reduce costs and encouraged greater clinician use of the tools, which would lead to higher patient enrollment and a higher likelihood that participating sites would sustain the program.

To combat issues with buy-in, sites used various methods to encourage clinicians to use the SMARTCare tools. In the third program year, one site developed a SMARTCare Clinicians On-Call program, which ensured that a clinician was always available to answer questions about the tools and provide additional guidance to patients on the merits of providing HIPAA authorization and participating in the program. Some sites held regular staff check-in meetings regarding SMARTCare implementation, which helped maintain engagement and encouraged continued use of the tools. Although sites employed this strategy from the beginning of the cooperative agreement, there was greater emphasis on these meetings in the third year. According to the survey results, the practice of having SMARTCare meetings seemed relatively widespread—88 percent of staff reported that they received information and instruction about the tools through staff meetings. Staff also learned about SMARTCare from team huddles (48 percent), mentoring (35 percent), individual supervision (32 percent), group supervision (27 percent), technical assistance (64 percent), and asking a colleague for help (70 percent). One interview respondent noted that the check-in meetings and on-call clinicians increased adherence from clinicians who were already participating in the program. This respondent hypothesized that clinicians are less likely to skip through the FOCUS tool when they are engaged through meetings. Another site encouraged tool use by giving out awards to clinicians who used the tools accurately or who received consent from the most patients.

Clinician engagement efforts were not successful at every site. Some clinicians either did not believe the tools were useful decision making aids or did not believe they needed decision making support. Interview respondents discussed how engagement strategies did not work well

with clinicians with this type of perception. One site dropped out of the program in the third year of the cooperative agreement, partially due to a lack of clinician engagement. Only one clinician at the site used the tools. Because of interoperability issues between the tools and the site's EMR system, manual work-arounds had to be put in place. Over time, the clinician felt overburdened with all the paperwork requirements and having to use a work-around to enter data into the tools. This clinician felt that the time commitment required by the program was too large to keep using the tools and enrolling patients, especially because the clinician was the only one participating at the site.

2. Delivery of program services

a. Description of and changes to service delivery model

With this cooperative agreement, the awardee set out to address a need they felt was important to incorporate while treating patients with SIHD. Before the cooperative agreement, clinicians had to make certain decisions about treating patients with SIHD, including whether to (1) order a stress test, (2) recommend invasive coronary angiography if stress test resulted in abnormal clinical results, and (3) recommend percutaneous coronary intervention (PCI) if the clinician performed an angiography. The awardee reported that clinical decisions did not always align with standards of appropriate use or take into account a patient's preference. Adding to this dilemma was the fact that there were geographical and institutional variations in the procedures used for SIHD as well as variations in the cost to patients. As a result, the awardee felt that there were clinical care gaps and disparities in care for appropriate use of procedures for SIHD patients.

Through the SMARTCare program, the awardee sought to integrate decision support into clinicians' workflow to reduce the inappropriate use of cardiac screening tests and procedures and to more heavily weigh patient preference in making treatment decisions. Specifically, the awardee did the following:

- Integrated a suite of clinical decision-support tools for cardiologists and other clinicians at the point of care
- Supported shared decision making with the patient
- Enabled the use of clinical registries to track and improve care

The awardee hypothesized that the program would result in improvements in the appropriate use of imaging tests and in patient care. The awardee also hypothesized that the use of the tools at participating sites would lead to a decrease in imaging tests and elective PCIs that did not meet the appropriate use criteria. In addition, through the shared decision making tools, the awardee anticipated improvements in both patient engagement and satisfaction. Last, the awardee anticipated that the overall program would lead to positive patient outcomes, including an increase in functional status, an increase in adherence to specific coronary artery disease treatment guidelines, and a decrease in the renal failure complication rate for elective cardiac invasive procedures and PCIs.

b. Evidence of service delivery effectiveness

Overall, the awardee was partially effective in delivering program services but was unable to do so as originally designed. Sites struggled throughout the cooperative agreement with integrating the program tools into their EMR systems and with initiating and maintaining clinician engagement. Sites also had uneven results in training the staff who were participating in the program. However, the sites paid close attention to the shared decision making component of the program and consistently updated resources to ensure that they met the needs of the patients.

Delivery of program services. The design of the program hinged on sites' ability to use the full suite of tools to fully deliver services as intended. Awardee leaders reported that one site that we did not interview implemented all tools as intended; however, none of the remaining sites were able to effectively do this, primarily because of health IT issues. Interoperability issues delayed services and led to variation in how sites used the tools, including using tools through a web-based system separate from the EMR and creating manual, paper-based work-arounds instead of using the FOCUS tool at the point of care.[3] None of the sites we interviewed implemented IndiGO, a key component of the service delivery model, because the sites could not integrate the tool within their EMRs.

Although the awardee was unable to deliver services as originally designed, both interview and survey respondents had generally positive feedback about some of the SMARTCare tools. However, they questioned the feasibility of working with multiple tools. One interview respondent noted that the tools were more efficient than previous procedures to address the appropriateness of test ordering but said that the tools would be more successful if they were integrated into one individual tool instead of separate and disparate tools. The FOCUS tool to assess patient risk appeared to be the most useful, as 70 percent of clinician survey respondents reported that the program impacted their ability to assess whether diagnostic tests for a patient met appropriate use criteria. Another interview respondent noted that Epic offered many of the same functions offered by IndiGO without the extra burden of needing to update versions of the tool and ensuring accessible functionality.

Staffing and training. Sites used existing staff to deliver program services. The awardee was partially effective in training individuals participating in the program. Although the awardee provided training to more than half of the clinicians (55 percent) and all of the staff (100 percent) surveyed, feedback about the training was mixed. Some respondents noted that they would have appreciated follow-up discussions or training after they became more familiar with the program and tools. Clinician training included instructions during staff meetings (76 percent), asking a colleague for help (56 percent), and self-study (43 percent). The modes of training were similar for program staff, who primarily received training through instructions during staff meetings (88 percent) but also learned to use the tools through self-study (83 percent) and asking a colleague for help (70 percent). Some of those clinicians who did not receive formal training (45 percent) stated that they would have liked some form of training.

[3] More details about the health IT barriers experienced in the first and second year of the cooperative agreement are available in the first (https://downloads.cms.gov/files/cmmi/hcia2-yroneevalrpt.pdf) and second (https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf) annual reports.

Recruitment and engagement of providers. The awardee was partially effective in engaging clinicians. The level of clinician engagement and willingness to participate in the program dictated how often the SMARTCare tools were used and program services were delivered to patients. Throughout the cooperative agreement, clinician engagement varied by site—mostly due to perceptions that the program offered clinicians minimal incentives to use the tool in a FFS environment and that clinicians did not believe they needed decision support. Discussions with site-level staff revealed that receptivity to the use of the SMARTCare tools varied not only across sites but also across potential users within the sites. Some clinicians bought into the program's goals, while others were skeptical of the proposed benefits. In Year 3, concern about the value of the program caused one site to stop participating and another site to consider dropping out.

Survey results support the idea that clinicians can be either a facilitator of or barrier to achieving the program's goals. Both clinicians (58 percent) and staff (63 percent) reported that resistance to the program by a clinician was a barrier. On the other hand, 46 percent of clinician respondents and 38 percent of staff respondents believed that clinician buy-in was helpful in achieving program goals.

Engagement of program participants. Shared decision making and functional status measurement were two of the key program features that aligned with engaging participants. The awardee has been effective in providing this component of the program. The awardee and its sites ensured that program services and education on shared decision making were meeting patient needs. In Year 1, the awardee responded to feedback about the ineffectiveness of patient education materials (for example, the availability of resources only in one language and requiring that patients return the informational DVD) by creating new shared decision making aids, such as a one-page overview of treatment options, to facilitate conversations between providers and patients. In addition, two of the sites in Year 3 piloted a new program workflow in order to provide the shared decision making resources and functional status baseline earlier in the process. The awardee found that having patients' survey results and treatment preferences up front enabled physicians to leverage functional status measurement as a clinical tool, in addition to an outcome measure, and to inject patients' needs and preferences into the care pathway earlier. The awardee reported that physicians implementing this workflow found it to be of value to their clinical decision making and saw an opportunity to further reduce tests and procedures in some cases. Survey results revealed that both clinicians (48 percent) and staff (57 percent) believed that they successfully engaged patients in the program.

c. Barriers and facilitators associated with service delivery effectiveness

> "EMRs that are interoperable and have data-extracting capabilities are needed to fully realize the effectiveness of SMARTCare's IT-based interventions."
>
> —ACCF respondent

In prior years, the awardee faced challenges with integrating the tools and incentivizing clinicians to use the tools. In the first year, the awardee quickly realized that implementing multiple tools across a variety of EMR platforms was a complex task that required coordination among different vendors and on-site technical support as well as a commitment from site leaders to ensure sufficient resources. Throughout the cooperative agreement, sites faced challenges in effectively integrating the tools. Not all sites were able to accomplish this goal because of interoperability

issues and incompatible security standards. Those sites that were successful noted that the service delivery model could not exist without health IT and emphasized how it helped facilitate the delivery of services.

The second barrier was the tension between reducing the amount of inappropriate imaging procedures ordered for patients and the lack of perceived clinician incentives to use the tool, in addition to lower procedure volumes that resulted in a reduction of payments to clinicians and health systems. Although the awardee believed that reducing the amount of inappropriate tests could result in better care for patients, clinicians and health care systems are not monetarily incentivized to reduce the number of tests they order in a FFS environment. For many of the clinicians interviewed, this was a major barrier for clinician buy-in for the program.

Staff who received training noted that training on the tools was a facilitator in their use. Fifty percent of clinicians and 100 percent of staff reported that they received training about the SMARTCare program. Of those staff members, 75 percent of staff and 71 percent of clinicians reported that they learned new skills that were important for their role with SMARTCare. Some staff and clinicians requested more training on patient consent and privacy and how other initiatives (such as the Medicare Access and CHIP Reauthorization Act) interact with the SMARTCare program and EMR dataflow.

C. Assessment of perceived program effects on the delivery of care and outcomes

Overall perceptions of the effects of SMARTCare on care delivery were mixed. Clinicians' survey feedback about the program revealed a positive outlook on the delivery of services but mixed opinions on the impact of the program on their work. More than 80 percent of respondents on the clinician survey reported that they thought SMARTCare was very or somewhat effective in achieving the program's stated goals. When asked about the impact of the program on the quality of care and services provided to patients, 59 percent of clinical respondents reported that the impact was positive. When asked whether they thought their ability to provide care or services to program participants was better, worse, or about the same as before the program began, 45 percent of surveyed clinicians said that the program improved their ability, while another 45 percent reported that their ability was about the same.

In interviews, respondents across sites believed that SMARTCare would achieve its goals of reducing inappropriate use of cardiac screening tests and procedures and incorporating patient preference in making treatment decisions. One respondent noted that site specific data showed a decreased in inappropriateness. However, some clinicians tempered this assessment by reporting that the current FFS environment in many practices makes the goal of reducing inappropriate testing difficult for physicians, who are a key driver for developing a payment methodology based on appropriate testing and outcomes. With regards to the shared decision making goal of the program, respondents reported that patients were more involved in their care through patient education, which was provided through the decision making aids and discussions with their providers. Although the awardee did not have publicly available data to share, one respondent said that patient survey results showed that shared decision making had increased at one site.

Other interview respondents did not believe that SMARTCare would meet its stated goals due to the timing of the cooperative agreement and a lack of physician engagement. First, sites had their program up and running at different times, mostly due to long contracting periods and issues with integrating the tools into their EMRs. As a result, some sites started enrolling patients later than planned and the number of enrolled patients varied across sites. Some sites felt that if they had more time, they would have enrolled more patients and would have been able to demonstrate the effectiveness of the program. Second, many sites had issues with creating and maintaining clinician engagement with the program. One interview respondent noted that the main reason for this was that there was no clear provider incentive to use the tools. In fact, the respondent felt that the tools encouraged less work. In addition, in a FFS environment, some clinicians did not see any value in participating. Respondents mentioned that other clinicians did not believe that there was a need for decision-support tools, and therefore the increased effort and paperwork was too much work. According to the survey, 26 percent of clinician respondents reported that SMARTCare increased their feelings of burnout "a lot" or "a little."

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Although some clinicians and staff perceived the SMARTCare program to have a positive impact, some implementation issues have implications for the interpretation and analysis of program impacts.

First, it may not be possible to demonstrate the desired outcomes because many clinicians chose not to use the FOCUS tool. In addition, if they did use the tool, many did not follow the decision support provided. Second, although the program included nine separate sites that may have followed the basic framework of the program's design, they also had individual differences and challenges that required specific tailoring of protocols and that impacted how the program operated. For example, one site was unable to integrate the FOCUS tool into its EMR, so staff had to create a paper-based work-around at the point of care. They would then input the data into the tool after seeing the patient. As a result of these perceived differences, interview respondents felt that it would be difficult to aggregate results across sites and generalize a finding on impacts. One respondent believes that sites would need to be assessed on an individual basis to better understand the program's impacts.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of the American College of Cardiology Foundation's SMARTCare program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: American College of Cardiology Foundation

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	2,454[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2016	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,716
Likelihood of all-cause hospitalizations	2,454
MDE sample size requirement to detect 20% effect	
Total expenditures	429
Likelihood of all-cause hospitalizations	614
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	Some issues, but probably surmountable. Expect to select a comparison group
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician and staff surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We will conduct a rigorous impact analysis of the SMARTCare intervention. Because the intervention is implemented by physician practices in Florida and Wisconsin, we plan to match participating physician practices to other practices in those two, or in adjacent, states that are similar in terms of provider specialty and practice size. We will use data provided by IMS Health for matching. From the matched practices, we will identify and select patients with indicators of stable ischemic heart disease to form the comparison group.

B. Characteristics of Medicare and Medicaid participants at baseline

Baseline data for the American College of Cardiology's SMARTCare program reflect a diversity of care settings but general uniformity in patient mix. This finding stems from the fact that the American College of Cardiology implemented the SMARTCare program in nine organizations in Wisconsin and Florida, including an academic medical center as well as nonprofit integrated health systems, hospital-owned outpatient clinics, and for-profit private practices. The group of participants represented both urban and rural settings. However, the program specifically targeted patients with SIHD who had indications for functional stress testing. Patients were enrolled in the program when their clinician completed, for their case, the FOCUS decision-support tool, which guided the appropriate ordering of cardiovascular imaging and tests.

We present data from all nine SMARTCare program sites. The 918 FFS Medicare beneficiaries represented in Tables III.2 to III.4 are only a small fraction of the 29,000 persons covered by private and public health plans who enrolled in the program and consented to sharing their data.[4] This report includes only the minority of SMARTCare participants who were Medicare FFS beneficiaries who consented to the use of their data. For the purposes of presenting baseline characteristics in this report, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who had met all program criteria for a period of 90 days within the 12-month period immediately prior to the start of their eligibility to receive program services. In addition, they were enrolled in the awardee's program on or before May 31, 2016, the most recent date for which we evaluated baseline characteristics. The calendar period covered by the baseline quarters is based on the enrollment date for each enrollee and therefore varies by enrollee. After we excluded patients who did not meet the above criteria, a total of 918 participants were included in the analysis of baseline characteristics for this report.

The Medicare FFS beneficiaries participating in the SMARTCare program in its first year of operation were mostly elderly, although few were among the oldest beneficiaries. Ten percent

[4] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

were under age 65, while 8 percent were over age 85. Slightly more than half (56 percent) were male, higher than the percentages of all Medicare beneficiaries in the two states (which range from 43 percent to 44 percent). By race, 85 percent were identified as white, 8 percent were black, and 6 percent were Hispanic or other. Twelve percent were dual eligibles, about the same as Medicare beneficiaries overall in these two states (Table III.2). At 1.1, the median hierarchical condition categories (HCC) score was only slightly above the national average of 1.0.

Participants in this initial group at baseline had somewhat higher monthly Medicare expenditures in the quarter prior to enrollment ($1,119) than in the year prior to enrollment ($986), perhaps reflecting recent conditions that precipitated cardiac diagnostic services. Both amounts were substantially higher than the average for the general Medicare population of $792 in 2014.[5] About one-quarter (23 percent) of the participants had a hospital admission in the year preceding enrollment. The participants averaged 7 primary care visits and 14 specialist visits in the preceding year (Table III.3).

Because the goal of the program was to reduce the percentage of imaging procedures not meeting the current appropriate use criteria, we report selected awardee-specific measures related to diagnosis and testing for cardiovascular conditions. The observed patterns in baseline claims reflect cardiovascular conditions expected in the target population. In the preceding year, 46 percent of participants had a diagnosis of chest pain, 95 percent had a cardiovascular diagnostic test, 81 percent had an electrocardiogram, and 49 percent had cardiovascular diagnostic imaging (Table III.4).

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the American College of Cardiology's program through May 31, 2016

	All participants (N = 918)	
Characteristics	Number	Percentage
Age as of enrollment date		
Younger than 65	96	10
65 to 74	463	50
75 to 84	290	32
85 and older	69	8
Gender		
Female	402	44
Male	516	56
Race		
White	783	85
Black	76	8

[5] For national average rates, see the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html.

Table III.2 *(continued)*

Characteristics	All participants (N = 918)	
	Number	Percentage
American Indian, Alaska Native, Asian/Pacific Island American, or other	31	3
Hispanic	12	1
Original reason for Medicare eligibility		
Old age and survivor's insurance	714	78
Disability insurance benefits	180	20
End-stage renal disease (ESRD)[a]	24	3
Hospice[b]		
Medicare/Medicaid dual status, percentage dual[b]	114	12
HCC score[c]		Statistic
Mean		1.48
25th percentile		0.72
Median		1.12
75th percentile		1.85

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which participating organizations report completing the FOCUS assessment tool for a beneficiary. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the American College of Cardiology's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	918	870	890	918	918
Average Medicare expenditures PBPM[a]					
Total	986 (50)	894 (75)	938 (101)	980 (79)	1,119 (82)
Acute inpatient	292 (28)	269 (44)	259 (85)	311 (53)	324 (49)
Inpatient other[b]	22 (7)	27 (16)	18 (11)	20 (15)	23 (12)
Outpatient[c]	237 (17)	236 (26)	241 (22)	203 (18)	269 (22)
Physician services	353 (15)	303 (25)	344 (21)	363 (22)	394 (17)
Home health	34 (5)	26 (6)	34 (8)	30 (7)	45 (10)
Skilled nursing facility	23 (6)	9 (5)	15 (7)	26 (12)	39 (16)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	25 (2)	24 (3)	26 (3)	27 (3)	26 (3)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	372 (29)	334 (45)	346 (55)	368 (47)	418 (50)
Outpatient ED visits	511 (37)	454 (63)	491 (55)	443 (51)	649 (65)
Observation stays	95 (12)	70 (19)	105 (21)	75 (24)	131 (24)
Primary care visits in any setting	6,625 (212)	5,929 (279)	6,091 (294)	6,206 (493)	8,179 (338)
Primary care visits in ambulatory settings	5,486 (154)	4,979 (190)	5,086 (214)	5,201 (460)	6,601 (217)
Specialist visits in any setting	13,741 (440)	12,178 (611)	13,074 (558)	12,878 (958)	16,662 (577)
Specialist visits in ambulatory settings	11,904 (349)	10,449 (456)	11,523 (425)	11,092 (896)	14,401 (423)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	23 (1)	7 (1)	7 (1)	8 (1)	9 (1)
Percentage with an outpatient ED visit[e]	32 (2)	9 (1)	10 (1)	9 (1)	13 (1)
Percentage with an observation stay[f]	8 (1)	2 (<0.5)	3 (1)	2 (<0.5)	3 (1)
Percentage with a 30-day readmission among all discharges	12 (2)	6 (3)	10 (4)	12 (4)	17 (4)
Percentage of participants with a readmission among all participants	3 (1)	0 (0)	1 (<0.5)	1 (<0.5)	1 (<0.5)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

Table III.4. Prevalence of chest pain diagnosis and prior service use among the American College of Cardiology's treatment group beneficiaries during the baseline period (N = 918)

Condition or service	Percentage
Chest pain diagnosis	46
Evaluation and management visit	72
Cardio diagnostic lab test	95
Electrocardiogram (ECG)	73
Diagnostic imaging	49
Noninvasive test	4
ECG monitoring	13

Source: Mathematica's analysis of information from the awardee's finder file and from Medicare claims and enrollment data as of May 31, 2016.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

The American College of Cardiology designed a condition-specific, bundled episode payment to support diagnosis and treatment of patients with SIHD. Services within the model include all evaluation and management (E&M) services by cardiologists between the time of the initial visit by the patient to a physician or cardiologist for new or significantly changed anginal symptoms and one month following the date when treatment for the symptoms was initiated. The bundled payment will also cover all stress tests, angiograms, and angioplasties with or without stenting received by the patient within the six months prior to the date when treatment for the current symptoms was initiated, regardless of what provider ordered or delivered the service.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

The bundled payment is triggered by a provider's use of the FOCUS tool for a patient in the ED or in an outpatient setting. The provider will use a specific E&M code to identify the patient. The hospital or outpatient practice will then submit a claim for the SMARTCare patient, which would indicate the patient's SMARTCare status. Upon receiving the claim, the payer will classify patients as either a Level 1–3 SMARTCARE patient or as a non-SMARTCare patient. The level of assignment is determined by (1) appropriateness of testing, (2) test results, and (3) quality of outcomes. If a cardiologist ordered a stress test for a patient when the FOCUS tool deemed it inappropriate but no other test or treatment were provided, then that patient would fall into the Level 1 category. On the other hand, if the cardiologist ordered a stress test when the FOCUS tool deemed it inappropriate and the test revealed that the patient needed bypass surgery, then that patient would be considered a Level 3.

Payers would provide reimbursement for services to the hospital or outpatient practice after a period of time predetermined by the payer and health care organizations. Before reimbursement, the payer will reconcile and determine whether the organization (hospital or outpatient practice) is over or under its projected budget for a group of patients. If it is over, the payer will bill the organization less. If it is under, then the organization gets to keep a portion of the savings as shared savings. A specific payment amount will be paid for each SMARTCare level. This amount will be determined by the payer and participating site (hospital or outpatient practice).

C. Status of the payment model

The awardee continues to refine the proposed bundled payment. The awardee is also working with an external payment model consultant. The awardee has not begun any negotiations but has begun conversations with various payers. Awardee leaders spoke to a small Wisconsin insurer to ensure that the bundled payment would be easy to administer and also to

identify other payment strategies. They have also held multiple conversations with a large national payer that expressed interest but wanted to work directly with sites and to focus on more narrow incentives, such as prior authorization waivers or contracts with quality measure performance bonuses.

The awardee realized from these conversations with insurers that the payment landscape is shifting so rapidly that large payers are less willing to invest in a payment model that has yet to have traction in the market. SMARTCare's bundle is perceived as being more complex than traditional pay-for-performance approaches. It is unlike other bundles in that it focuses on the diagnosis, disease prevention, and treatment processes. In addition, because the risk stratification in the model is based on clinical data not currently collected in claims, payers are unable to do their own modeling of impacts.

The awardee does not anticipate implementation of the model before the close of the cooperative agreement. Next steps in development and finalization of the payment model include the following:

- Collection of data from sites on the volumes and cost of services
- Completion of the payment modeling exercise using manually collected data from one site
- Revision of the payment model based on modeling and feedback
- Exploration of a payment strategy to support shared decision making and functional status measurement

D. Factors associated with the development of the payment model

The awardee noted two key factors influencing the development of the payment model: (1) feedback from business leaders and (2) feedback from site-level staff. Before the cooperative agreement began, the awardee started initial discussions with the business consortium of Wisconsin to help develop the payment model. The awardee said that the business consortium's feedback to simplify the payment model and feedback from other stakeholders to add different payment levels has been crucial to moving the payment model forward. In addition, the awardee said that engaging pilot site financial leaders has been enormously instructive. Given the complexity and rapid change characterizing the current payment landscape, site financial leaders helped awardee team members understand the important nuances of their markets and signaled that they needed to be more fully engaged in understanding and promoting the SMARTCare model as a part of their financial strategy.

The biggest challenge to developing the payment model was difficulty in obtaining data. The awardee has been working with one site on manual collection of data in order to test the model by using real site data. However, this has been a large undertaking because the process is manual and requires line-by-line patient matching in Excel. Although one site was willing to do this for the project, this is not sustainable or replicable. Obtaining access to patient matched data contained within the SMARTCare database was cost prohibitive and not included in the grant. Concerns for data safety and security and the lack of interoperability of billing systems with clinical data limited the ability to easily link the data.

Another challenge has been payer engagement. The awardee reported that payers it approached seemed less interested in payment reform than the awardee expected. As mentioned in the previous section, the awardee believes payers are less willing to invest in an untested payment model, especially with a changing payment landscape that they believe is moving away from bundles. Self-insured employers have expressed more interest, but their small individual market share makes it difficult to assemble a critical mass of payments to support the program.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The American College of Cardiology reported that almost all implementing sites will sustain some aspects of the SMARTCare program, as of the third program year. The awardee created program resources that implementing sites can use to sustain aspects of the program most valuable to them. The awardee did not report plans to scale or replicate the program, although it noted that program resources can be used to replicate the program in the future.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, the American College of Cardiology reported plans to sustain the program by simplifying it, training implementing sites to operate the program independently, and developing the payment model. To simplify the program, the awardee planned to integrate the program's clinical decision support, shared decision making, and health IT program components. The awardee was training sites to operate the program without support from the awardee through one-on-one coaching and mentoring. The awardee was also developing and disseminating written guidance on self-maintenance.

C. Implementing the SSR plan: progress and changes

Sustainability. At the end of the third program year, the awardee reported that eight of the nine implementing sites would sustain at least one aspect of the program. None of the site staff we interviewed reported certainty that the whole program would be sustained at their site. Aspects of the program that will be sustained include ePRISM, clinical decision support, and the program's general impact on the site's strategic planning. One site reported that, although most of the program would not be sustained, it planned to incorporate similar features from the program tools into its EMR. The awardee also reported creating program resources to help sites sustain aspects of the program they find valuable, such as a video explaining the program and a "SMARTCare 101" fact sheet. These resources (available at http://www.e4enterprise.com/SMARTCare/Playbook/) are organized by the four core components of SMARTCare, focus on how to adapt the program components to evolving payment and delivery system models, and allow sites to pick the components most valuable to them. The awardee thought these resources may help replicate the program in the future, although neither the awardee nor the implementing sites reported concrete plans to do so by the end of the third program year.

Scalability. The American College of Cardiology did not report plans to scale its program in the third program year.

Replicability. The American College of Cardiology did not report plans to replicate its program in the third program year.

D. Factors associated with progress toward implementing the SSR plan

Only one out of nine sites will continue to use the FOCUS tool. None of the other main tools within the program will be sustained by any of the sites, according to a quarterly report submitted by the awardee to the implementation and monitoring contractor. Program cost and competing organizational priorities were the two main challenges to sustainment reported by the awardee. Multiple respondents, including awardee and site leaders, reported that the cost of the health IT tools prohibited sustainment. The biggest challenge was funding the constant maintenance required for properly functioning tools. In addition to having to pay for licensing fees, the tools must be regularly updated to remain interoperable with the EMRs. Moreover, during the awardee's final round of site visits to the implementing sites at the end of the third program year, the awardee heard from sites that their organizations could not prioritize the program. Instead, they were focused on other sources of revenue as a result of the largely FFS environment in which they operate. Regardless, the awardee expressed belief that implementing sites will be able to sustain aspects of the program, given program resources developed by the awardee.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension to capture lessons learned, develop case summaries, and conduct project reporting. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in the American College of Cardiology's SMARTCare program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicare beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Altarum

April 18, 2018

Dana Petersen (Mathematica Policy Research)
Theresa Feeley-Summerl (Mathematica Policy Research)
Kristin Andrews Lemos (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Altarum received a six-month extension through February 28, 2018, and provided some services until the end of the extension.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

Dental disease is one of the most prevalent chronic conditions in childhood. It is a persistent problem for the youngest children (birth to 3 years of age), particularly those from families with low incomes.[4] Altarum, a nonprofit health research organization, partnered with Delta Dental of Michigan, the University of Michigan (UM) School of Dentistry, and the Michigan Department of Health and Human Services (DHHS), the state Medicaid agency, to use funding from HCIA R2 to implement the Michigan Caries Prevention Program (MCPP) (Table I.1). The MCPP was designed to fill a critical care gap in preventive dental care in early childhood, and to encourage the establishment of dental homes earlier in childhood than has historically occurred for Medicaid and Children's Health Insurance Program (CHIP) beneficiaries in the state. The program's goals were to (1) expand delivery of preventive oral health services by primary care providers (PCPs) who serve children, (2) integrate oral health care across primary care and dental settings, and (3) build the capacity of the oral health safety net.

The MCPP was innovative in its multifaceted approach to building multiple audiences' (PCPs, dentists, public health professionals, educators, caregivers, and children) awareness of the importance of dental care in early childhood and enhancing their ability to improve the dental health of the youngest children. Many programs designed to improve oral health focus on only one audience instead of integrating approaches and building linkages across medical, dental, community, and family systems.

Key MCPP components included (1) training and technical assistance (TA) given to PCPs and their office staff to build their capacity to deliver evidence-based preventive oral health services and refer patients to dental homes, (2) a health information technology (health IT) system to facilitate communication and help coordinate referrals between medical and dental providers, and (3) participant and family engagement, including educational outreach to dentists, dental hygienists, and public health professionals; training for staff at Women, Infants, and Children (WIC) clinics; the use of broad-based dissemination strategies such as conference presentations and social media; and a crowdsourcing website (SmileConnect) to help fill needs for early dental health services and supplies.

[4] Clark, M. B., and R .L. Slayton. "Fluoride Use in Caries Prevention in the Primary Care Setting." *Pediatrics,* vol. 134, no. 3, 2014, pp. 626–633.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The MCPP was designed to improve access to preventive dental care and encourage the establishment of dental homes for the youngest Medicaid and CHIP beneficiaries in the state.
Major innovation	The MCPP was innovative because it sought to improve access to and delivery of preventive oral health care through a variety of audiences. The program used several strategies to increase the capacity of PCPs, dental providers, public health professionals, educators, caregivers, and children to promote children's oral health.
Program components	Education and training Health IT Participant and family engagement
Target population	742,715 Medicaid and CHIP beneficiaries in Michigan who are age 17 or younger.[a]
Theory of change/ theory of action	Altarum hypothesized that the program would improve access to dental care by providing care in primary health care and community settings and increasing referrals from primary care providers to dentists. Altarum hypothesized that if more children received preventive oral health care, it would reduce the amount of dental disease, the number of dental caries, and the costs associated with untreated dental caries.
Payment model	New fee-for-service (FFS) payment, value-based payments
Award amount	$9,383,762
Effective launch date	May 8, 2015
Program setting	Pediatric provider clinics and medical centers; dental offices; and other community, early education, school, and public health settings
Market area	Urban and rural
Market location	Michigan
Target outcomes	• Increase the proportion of children in low-income households who receive preventive oral health care by 60 percent • Reduce the proportion of Medicaid beneficiaries who have dental caries by 30 percent • Provide a net savings to CMS of $21.1 million

[a]Altarum lowered its enrollment goal from just over 1 million indirect participants in the third program year.

CHIP = Children's Health Insurance Program; IT = information technology; MCPP = Michigan Caries Prevention Program.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that Altarum was partly successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, the awardee enrolled 949,164 indirect participants—128 percent of its enrollment target—by the end of the initial cooperative agreement. Second, Altarum and its partners implemented the three key components of their model mostly as planned. For example, they delivered preventive oral health training and TA as planned to more than 3,000 PCPs and clinical support staff; launched Michigan's Dental Registry (MiDR), a public health improvement and health IT tool, after significant delays in developing it; conducted outreach and education via WIC programming; and created a crowdsourcing website to connect communities with oral health needs to dental

providers who could fulfill them. Third, Altarum recruited, hired, and retained the staff it needed to develop and deliver all MCPP components. Fourth, while the awardee was successful in engaging PCPs and WIC staff in training and TA activities, it faced significant challenges with directly engaging dental providers, promoting their use of MiDR, and fostering communication and linkages between medical and dental settings. Finally, awardee leaders, staff, and providers who participated in training activities thought the program had positive effects on the delivery of oral health care.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Altarum's MCPP program, the analysis is still in progress and not included in this report.

Payment model. By the end of the third program year, Altarum had obtained Medicaid claims data and started using the data to develop a formal value model (that is, evaluate the potential benefits and costs of its program in order to justify the cost of the financial incentives proposed in the payment model). Altarum's proposed payment model is an enhanced fee-for-service (FFS) model with the potential for bonus payments. It features the following elements: (1) higher reimbursement rates for both preventive oral health services and dental treatment, and (2) a bonus payment to dental providers who reach a yet-to-be determined threshold (that is, a percentage) of their pediatric patients receiving dental preventive care.

Sustainability plans. Altarum secured $500,000 in annual state funding to maintain the health IT system. State policy changes beginning in 2018 also required broader adoption of the health IT system among dental providers who receive Medicaid reimbursement. The awardee is also working with the University of California Los Angeles (UCLA) to replicate and enhance the training and education and health IT components of the program in Los Angeles (LA) County. However, Altarum had not had significant success in securing funding to sustain the training and TA functions in Michigan. The awardee also planned to maintain and expand its crowdsourcing website and decided to brand the entire MCPP program as SmileConnect to signal the interconnectedness of all of the program's components and its transferability to other states.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The clinician survey was fielded from March to June of 2017 with 300 clinicians who participated in MCPP training and achieved a response rate of 62 percent. The non-clinician staff survey was fielded from July to October of 2016 with a sample of 99 non-clinical office staff who also participated in MCPP training and that achieved a response rate of 61 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs and/or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that Altarum was partly successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

Altarum initially sought to train any PCP who volunteered to participate in its program. The awardee recruited PCPs through relationships that Altarum and its partners had established with primary care practices serving children in other projects and through broad-based efforts, such as mailing flyers to pediatric practices that could reach a wide audience. In the second program year, the awardee changed the focus of its recruiting efforts and began proactively targeting PCPs who were likely to serve large Medicaid populations, such as those working at federally qualified health centers (FQHCs). They also focused recruitment on residency programs. The awardee also began recruiting rural health centers in the third program year. In May 2017,

Altarum stopped provider recruitment efforts so that it could finish training the enrolled practices before the award ended.

The MCPP does not provide services directly to patients. Rather, the program trains PCPs who treat the target population and considers all children served by sites with trained PCPs as indirect participants. When providers sign up for MCPP training, they estimate the number of patients at their site who are 17 years old and younger. These figures become the basis for estimating the number of indirect participants served. Participants indirectly served through SmileConnect, the awardee's crowdsourcing website, are also included in the count of indirect participants served.

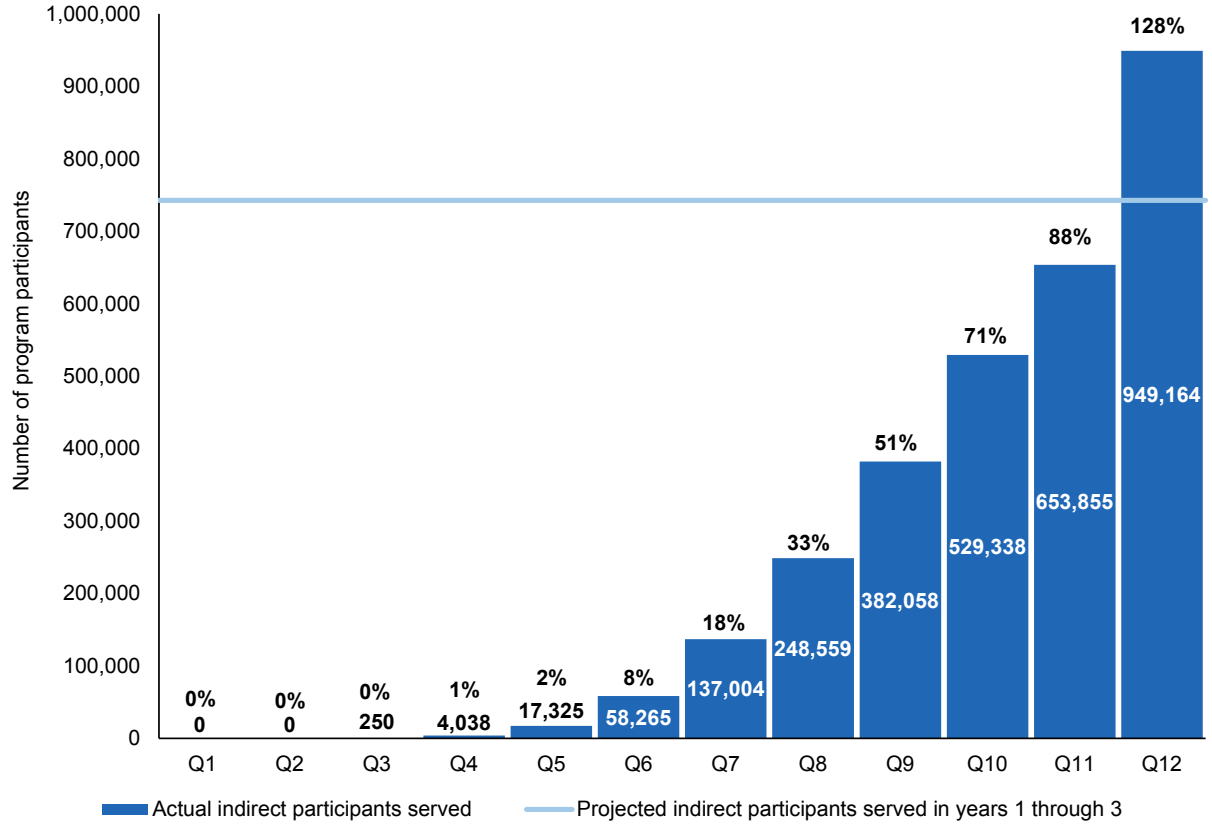
b. Evidence of enrollment effectiveness

Altarum achieved enrollment effectiveness. The awardee trained 1,565 PCPs from May 2015 (when it launched its program) through August 2017. This represents 104 percent of its three-year enrollment target for providers. The awardee also trained 1,588 clinical support staff in the participating practices.

After a slow start enrolling participants during the first six program quarters, Altarum began making steady progress and experienced a surge in participant enrollment in the final quarter. As a result, the awardee surpassed its enrollment goal for indirect participants. Throughout the three program years, the MCPP indirectly served 949,164 participants from May 2015 (when it launched its program) through August 2017, which represents 128 percent of its projected number of indirect participants (Figure II.1). This count includes participants served by providers trained through the MCPP as well as those who received services and supplies through SmileConnect.

Altarum lowered its enrollment target from about 1 million Medicaid and CHIP participants to 742,715 during the third program year because fewer indirect participants were served by medical residents than expected. However, Altarum continued to target medical residents for MCPP training because the awardee believed residents would use evidence-based oral health care practices in the clinics where they would ultimately be practicing.

Figure II.1. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers. Altarum lowered its projected number of indirect participants from roughly 1 million to 742,715 in the third program year. Altarum does not have direct participants.

c. Barriers and facilitators associated with enrollment effectiveness

Altarum's progress in meeting its three-year enrollment goal for indirect participants was facilitated by two factors: (1) the availability of provider participation incentives for PCPs and (2) peer-to-peer endorsement of the program. Two factors that kept more indirect participants from being enrolled were (1) the type of providers who enrolled and (2) the challenges of offering services in rural areas of Michigan.

In interviews and reports, Altarum leaders repeatedly told us that the most important factor driving providers to enroll in the training and TA component was the provision of continuing medical education (CME) and maintenance of certification (MOC) Part IV credits to PCPs who completed the training. These credits were thus a prime facilitator in enrolling indirect participants. In the first program year, Altarum built the jointly sponsored CME and MOC Part IV program with the UM Medical School and Health System to encourage participation in the

training and increase the likelihood of sustained change in the practices.[5] Further incentivizing providers' participation, the American Medical Association (AMA) increased the number of CME and MOC Part IV credits allocated to training participants because of the high quality of MCPP's quality improvement (QI) activities in the sixth program quarter.

Awardee leaders said that word-of-mouth promotion of the program by past trainees to potential trainees spurred provider participation (and thus resulted in more indirect participants). The leaders said that some health systems were increasingly willing to enroll more of their clinics after their residency programs had undergone MCPP training and they had witnessed the results. Finally, awardee leaders described numerous ways that they leveraged the connections of the leadership team and the MCPP Physician Advisory Committee to endorse the MCPP and recruit PCPs. For example, one advisory committee member encouraged FQHCs to participate by relying on his network of chief medical officers at the health centers.

> "We have a lot of physician organizations that had heard really good things from the physicians who participated and were endorsing the program. That gave us legitimacy, and clinics started reaching out to us."
>
> —Altarum leader

Altarum faced its main barrier to enrolling indirect participants during the first program year. Many of the PCPs who volunteered for and received training were in suburban practices, and there were few Medicaid beneficiaries in their service areas. To overcome this barrier and boost the number of Medicaid participants indirectly served, the awardee refocused its recruitment strategies in summer 2015 and began focusing on PCPs and health systems that were likely to serve large Medicaid populations, such as FQHCs.

Altarum leaders also reported that it was difficult to find qualified implementation specialists (Altarum staff who provide the Smiles for Life training to clinics) to offer PCP training in remote areas of Michigan such as the Upper Peninsula. Because the awardee did not hire staff in remote areas, it could not recruit providers or provide in-person training there, and thus Altarum did not enroll participants in these areas. In the second program year, Altarum's ability to offer CME and MOC Part IV credits was extended to cover live-streamed training. At the beginning of the third program year, Altarum began recruiting providers and offering online training in remote areas (that is, live-streamed training). Consequently, more indirect participants enrolled in the program. In awardee reports, Altarum described similar difficulties in finding dental providers who could fulfill SmileConnect requests (and thus indirectly serve participants) in rural areas. In interviews during summer 2017, an awardee leader reported trying to overcome this challenge by cold-calling dental providers in these areas to encourage their participation, or asking dental providers in more urban areas to travel to remote settings to provide services.

[5] To earn CME or MOC credits, participating providers have to demonstrate that they "meaningfully participated" in the performance improvement activity—in this case, the MCPP interventions. A physician at the University of Michigan Medical School, in partnership with Altarum, attests to whether or not providers have meaningfully participated based on improvements in their provision of the targeted oral health care services from baseline through the seven-month follow-up period.

2. Delivery of program services

a. Description of and changes to service delivery model

The MCPP was a new project initiated by Altarum and its partners under the HCIA R2 cooperative agreement. Although each partner brought relevant experience and expertise, the three organizations had not worked together before, nor had the program services existed in Michigan before.

Altarum and its partners sought to improve PCPs', dental providers', and public health providers' ability to deliver evidence-based oral health care to young children through three intervention components that did not change significantly during the award. In doing so, they hypothesized that children would receive preventive oral health care and education at a younger age, adopt positive oral health behaviors, experience fewer dental caries and less dental disease, and thus have lower spending on oral health care. The awardee also expected use of its public health improvement and health IT tool (MiDR[6]) to increase coordination and communication between medical and dental providers and thus positively affect the quality of early oral health care. In Table II.2, we describe the processes that Alarum believed were required to achieve desired outcomes for each MCPP component.

Table II.2. Key service delivery activities

Components	Key activities
Training and TA	• Identifying a motivated individual to serve as the "oral health champion" throughout the training period to help Altarum tailor the training to the site's specific needs, serve as the liaison between the site and Altarum, and promote the training among colleagues • Providing a one-time clinic-wide training to PCPs serving children and their clinical office staff that covered evidence-based standards of oral health care using the Smiles for Life[a] curriculum • Providing ongoing TA for up to seven months after the initial training for sites enrolled in the CME and MOC Part IV performance and QI activity and for up to four months for other sites. • Having PCPs adopt and hone their skills in evidence-based oral health practices (that is, oral health screening, application of fluoride varnish, oral health education, referral to dental homes) and integrate the practices into their usual care delivery • Having PCPs make referrals to dentists and coordinate with dental providers to establish dental homes for children
Health IT	• Developing a public health improvement and health IT system (MiDR) for medical and dental providers to document oral health services provided to children and refer patients to dentists and dental specialists accepting Medicaid • Classifying MiDR as a Meaningful Use Specialized Registry that receives data from medical and dental providers on the oral health care they provide • Training medical and dental providers to use MiDR • Having medical and dental providers enter oral health services provided to children into MiDR and use the system to make referrals and coordinate children's oral health care • Having the state use MiDR to track measures of the quality of children's dental health

[6] MiDR includes a web-based and an electronic medical record interface to facilitate documentation of preventive oral health service provision and referrals to dental providers as well as coordination between primary care and dental providers. The EMR module also provides decision support and includes a screening tool for oral health risks. MiDR is classified as a Meaningful Use Specialized Registry.

Table II.2 *(continued)*

Components	Key activities
Participant and family engagement	• Engaging in large-scale dissemination to and education of dental providers, via conferences and presentations, about the need to provide care to young children and provided tips on providing care to young children • Training Women, Infants, and Children (WIC) staff through the WIC BRUSH curriculum to integrate oral health counseling into their programs • Having WIC staff provide oral health education to the children and families they serve and refer children to dental providers • Creating a crowdsourcing website (SmileConnect) to connect (1) users such as community organizations and early childhood educators who post requests for oral health services or supplies with (2) users such as dental providers and philanthropic organizations who can fulfill those requests • Using social media—including an MCPP website and Twitter— to build public awareness of the importance of oral health care for children and to create demand for oral health care

[a]Smiles for Life is a national oral health curriculum. Since 2008, Michigan Medicaid has reimbursed pediatric primary care providers (including physicians and nurse practitioners) who receive training and certification for conducting oral health screening and applying fluoride varnish for children up to 35 months of age. In July 2012, the state required Smiles for Life certification for reimbursement for these services. The Smiles for Life editor gave permission to Altarum to use the curriculum and to directly certify physicians and nurse practitioners who complete the MCPP training in Smiles for Life.

b. Evidence of service delivery effectiveness

Altarum partly achieved service delivery effectiveness. The awardee was effective in delivering its training and TA to medical providers as planned and on schedule. It was also effective in delivering its family and patient engagement service component, especially through its WIC programming and SmileConnect crowdsourcing website. In addition, the awardee successfully recruited, hired, and retained program staff and engaged PCPs and clinical office staff in its training and TA component. However, the awardee and its partners experienced significant delays in developing MiDR and encountered significant challenges in engaging medical and especially dental providers to use the MiDR system. These challenges hindered the awardee's ability to create the linkages it had intended between medical and dental providers. Details follow below.

Delivery of intervention services. According to awardee leaders and staff, Altarum and its partners were successful in developing and delivering training and TA. The awardee adapted the Smiles for Life curriculum for use in the MCPP and began offering in-person provider training as planned during the first program year. It started offering an online version of the training during the third program year. The awardee reported that, as of August 2017, it had operated 23 cycles of training totaling 2,717 hours of instruction. The awardee reported training at least one-third of all PCPs serving children covered by Medicaid in the state, and said that physicians at more than 150 practices have received CME or MOC Part IV credit. In interviews, implementation specialists reported delivering scheduled follow-up TA as planned one month after the initial training and every three months thereafter, and on an ad hoc basis in response to clinic requests or performance on monitoring measures. Clinicians who participated in the MCPP training and responded to the clinician survey during the second half of the third program year confirmed that they had received training in the application of fluoride varnish (99 percent), oral health screening (94 percent), patient and family engagement on oral health care (92 percent), and referrals to dentists (79 percent)—all core topics areas of the curriculum. With regard to the

quality of training provided by MCPP, nearly all clinicians (93 percent) strongly or somewhat agreed that they learned new skills that are important for their roles, and 89 percent strongly or somewhat agreed that the training helped them improve their job performance.

Although Altarum was successful in developing MiDR, the awardee experienced significant delays along the way. For example, Altarum and its partners launched a pilot test of MiDR's electronic medical record (EMR) module and functionality in one practice in June 2016, about one year later than originally planned, and the web-based version of the system did not go live until October 2016. The awardee also integrated MiDR with Michigan's single web portal, which state employees and providers who treat children already use to access multiple state applications. Altarum collaborated with a health IT vendor to integrate MiDR with an EMR commonly used in the state. Altarum then joined a private sector initiative composed of several EMR stakeholders, and was able to integrate MiDR into more EMRs. The awardee reported that as of August 2017, MiDR had 500 active users (two-thirds of users were medical providers, and one-third were dental providers) and that more than 3,600 treatment updates and 400 referrals had been documented in the system. Delays in developing the MiDR (described below) meant that the awardee rolled out training and TA without an operational health IT system—in contrast with its original plans. In interviews conducted in summer 2017, staff reported reaching out to all clinics trained before MiDR launched to alert them of MiDR's availability. However, staff were only able to reach about half of the clinics that were trained before the MiDR launch to give them additional training on MiDR.

Finally, Altarum successfully implemented activities to engage patients and families—and did so on a scope wider than it originally planned. For example, the awardee and its partners leveraged outside funding to train WIC staff to support oral health among the families they serve at 16 sites, twice as many as planned. Altarum was also successful in launching the SmileConnect website on schedule in February 2016 and launching a mobile application in the third program year. The awardee reported making numerous conference presentations and using ongoing social media strategies to promote dental providers' use of evidence-based early oral health care practices and their engagement with SmileConnect. The awardee reported that it had fulfilled nearly 200 SmileConnect requests for dental health services or supplies by the end of the 12th program quarter.

Staffing and training. Altarum was successful in recruiting, hiring, and retaining the staff it needed to develop and deliver MCPP components. The awardee reported that as of August 2017, it had hired 24.5 full-time equivalent staff, which was 153 percent of its three-year hiring projection, and had achieved a 97 percent staff retention rate. Altarum's staff consisted of about 30 people in various roles, who carried out the demands of the program components. Most staff were located at Altarum's headquarters. Implementation specialists were geographically dispersed so they could recruit sites and provide training and TA locally. Altarum also said it effectively leveraged its partners' staff to implement MCPP program components. For example, Delta Dental staff participated in MiDR development and outreach to dentists, and UM School of Dentistry staff adapted Smiles for Life training materials, trained implementation specialists, and identified providers to fulfill SmileConnect requests. Michigan DHHS staff supported MiDR development and the awardee's efforts to engage PCPs, dental providers, and participants and

families. Awardee leaders described significant staff turnover at Michigan DHHS, but not among other partners.

In interviews in summer 2017, staff reported receiving adequate training or other support to perform their duties. For the most part, Altarum hired people who already had the requisite expertise, and therefore the awardee did not provide formal training. Awardee leaders and implementation specialists described the training provided by the UM School of Dentistry staff on the Smiles for Life curriculum as high in quality, and especially appreciated the observations they conducted of each staff's first training to provide coaching and feedback before clearing staff to provide training independently. Staff also appreciated informal training they received through monthly implementation data meetings. At these meetings, staff reviewed participating sites' performance data and collectively brainstormed solutions to the challenges practices were encountering with awardee leaders and other implementation specialists.

Recruitment and engagement of providers. Awardee leader, staff, and partner assessments of provider engagement varied depending on the MCPP component. According to staff in interviews, for example, the majority of providers and clinical support staff participated actively in initial trainings, and Altarum was able to rely on sites' oral health champions, as intended, to encourage and sustain its engagement of providers over time. Staff and partners also said that WIC staff who participated in WIC BRUSH trainings were highly motivated and engaged, yet they also said it was difficult to engage early childhood educators to make requests—and to get dental providers to fulfill requests—via SmileConnect in the early years of the cooperative agreement. In addition, awardee leaders and partners described ongoing struggles to achieve the desired level of use of MiDR among providers, especially dental providers.

> "In the medical office, the culture shift has already happened around technology adoption, so we're now trying to encourage the dental office to adopt technology. It's a challenge because we're starting from square one."
>
> —Altarum leader

Engagement of program participants. Altarum did not include direct engagement of program participants in its three primary program components. Instead, it engaged indirectly with participants through intervention components.

c. Barriers and facilitators associated with service delivery effectiveness

Altarum's ability to effectively deliver intervention components was influenced by several factors. Looking over the full three-year cooperative agreement, Altarum identified four primary facilitators to service delivery: (1) early and ongoing outreach to stakeholders, (2) the QI requirements of the CME and MOC Part IV program, (3) the staffing model that incorporated implementation specialists, and (4) experience with health IT development. MCPP leaders also described two primary barriers to the delivery of intervention components: (1) organizational changes at Michigan DHHS, and (2) limited engagement of dental providers.

First, Altarum leaders and leaders at its partners noted that bringing in stakeholders early and working to keep them engaged made it easier to implement MCPP components. For example, Altarum's Physician Advisory Committee—which included representatives from the Michigan Oral Health Coalition, the Michigan Primary Care Association, and other organizations—disseminated information about the importance of early dental care and encouraged their constituents to participate in MiDR. In addition, the America's Tooth Fairy organization offered logistic support to the Altarum team in fulfilling SmileConnect requests for dental services and supplies.

> "Looking to the stakeholder engagement piece early on [was helpful. We conducted an] environmental scan of the state and looked to align and engage with other groups that were on existing initiatives. We incorporated them into the planning process so they felt included, and they became allies as opposed to roadblocks later."
>
> —Altarum leader

Second, Altarum leaders said that the QI requirements of the CME and MOC Part IV program supported their ability to provide high-quality training and TA that was tailored to the specific needs of participating practices. The program required participating PCPs to submit monitoring data at baseline and at four and seven months post-training. Staff described reviewing data submitted at four months to identify areas for targeted TA to practices and analyzing data submitted at seven months identify facilitators and barriers and to practices' success and refine their approach for future trainings.

Third, Altarum leaders and staff at participating practices cited aspects of the staffing model that incorporated implementation specialists as a facilitator of the MCPP training and TA component. According to leaders, hiring geographically dispersed implementation specialist staff enabled them to gain familiarity with local context and to be available to provide hands-on support when needed. Altarum leaders also said that, when hiring staff, prioritizing QI backgrounds and experience in clinical workflow over dental health expertise accelerated sites' ability to incorporate evidence-based oral health practices into well child visits. In summer 2017 interviews, providers who participated in the training corroborated these practices as facilitating factors.

> "Our [implementation] specialist with Altarum has been great. If I need something, I'll just call her or shoot her an email. She's always willing to either come down and help me or do it over the phone."
>
> —Non-clinical staff at participating practice

Fourth, Altarum said its experience in developing health IT helped as the awardee built MiDR. For example, Altarum's developers learned lessons about the requirements for a state-run system while building the Michigan Care Improvement Registry (MCIR),[7] and they applied those lessons to the design of MiDR. This helped Altarum integrate MiDR with Michigan's single web sign-on portal.

A significant barrier to Altarum's effectiveness in developing MiDR on schedule was the reorganization of the Michigan DHHS after the awardee submitted the HCIA R2 proposal.

[7] The MCIR collects immunization information for children and is accessible to authorized users, such as health care providers. It provides up-to-date immunization information for health care organizations, schools, child care programs, pharmacies, and Michigan residents.

Because several people at DHHS who were involved in writing the proposal were no longer employed by the state at the time of the award, the awardee spent more time and resources than anticipated to identify DHHS staff who would champion the work on MiDR. Through its persistence in engaging stakeholders within DHHS, its success in establishing MiDR as a Meaningful Use Specialized Registry with capabilities to support statewide quality monitoring, and its receipt of the HCIA R2 funds that supported the work, the awardee eventually convinced DHHS to prioritize the development of MiDR and its integration with state systems.

The awardee and its partners also singled out the historic lack of coordination between PCPs and dentists and the dental community's limited experience with health IT and quality performance requirements as two persistent barriers to the implementation's effectiveness. These barriers kept more dental providers from participating with MiDR and hindered the awardee's efforts to promote coordination and referrals between PCPs and dentists.

C. Assessment of perceived program effects on the delivery of care and outcomes

All awardee leaders and staff interviewed in summer 2017 believed the MCPP had a positive effect on the delivery of oral health care by PCP and WIC providers. They were less certain about the program's effect on the delivery of oral health care by dentists.

> "I think [the potential impact of the program is] huge just because we're always missing that education factor. In the past we could tell them, "Yes, brush your teeth," but we didn't tell them why. We're able to do that now."
>
> —Provider at participating clinic

In interviews, awardee leaders and staff agreed that PCPs had improved their delivery of oral health services, crediting them with providing preventive services, counseling participants and families on the importance of oral health, and referring the children and their families to appropriate dental care. They also routinely cited CME and MOC Part IV monitoring data as evidence of the training and TA component's lasting positive influence on PCPs' behavior and on service delivery. In its 12th quarterly report to CMS, the awardee reported that across 159 practices trained, the weighted rates of oral health screening, fluoride varnish applications, and dental home referrals increased by 81 percentage points, 68 percentage points, and 60 percentage points, respectively, between baseline and five to six months post-training.

PCPs and clinical staff who participated in the MCPP training mostly corroborated the awardee's viewpoints. The majority of respondents to the clinician and non-clinician surveys believed the MCPP training had positive impacts on the delivery of care and on participant and family outcomes (Table II.3). In addition, nearly all PCPs who completed the clinician survey indicated that the program was very (37 percent) or somewhat (59 percent) effective in achieving its goals. Contradicting CME and MOC Part IV data, however, only 30 percent of PCPs surveyed during the second half of the third program year reported that, compared with a typical week before they

> "We were able to significantly impact provider behavior, training at least one-third of the primary care physicians serving Michigan Medicaid patients. We've seen an increase in preventive oral health services (fluoride varnish and oral screenings), and developed a referral management system that has become mandated for dental Medicaid providers to use."
>
> —Altarum report to CMS, quarter 12

started participating in MCPP, they spent more time referring participants to dentists and following up on referrals. Only 10 percent said they spent more time coordinating care and communicating with dentists.

Table II.3. Staff and clinicians' perceptions of the MCPP's effects on service delivery

The MCPP had a positive impact on:	Percent of respondents who agreed with the statement	
	Clinicians trained (n = 167)	Non-clinician staff trained (n = 40)
The quality of care and services you provide to participants	90	90
Your ability to respond in a timely way to participant needs	49	63
Your ability to provide care or services that are responsive to participant preferences, needs, and values	72	82
Care services that are provided fairly to all participants	N/A	79
Access to care or services for all participants	81	87
Achievement of participants' health goals	88	90
Participant satisfaction	80	87
Participant quality of life	82	84
Care coordination	64	87

Source: HCIA R2 evaluation survey of awardee's non-clinician staff, July to October 2016 and clinician staff survey, March to June of 2017. A few responses were missing for some survey clinician survey items (one to four) and non-clinician survey items (one or two).

N/A = not available.

Awardee leaders and staff also believed that the WIC BRUSH training program was effective. They said that WIC staff had integrated oral health education and referrals into their workflows, and the program positively affected child and family oral health behavior and access to dental care. In its 12th quarterly report to CMS, the awardee reported initial results from an internal analysis of data from first five WIC sites and 30 WIC staff trained by MCPP. According to the awardee, children served at WIC BRUSH sites were more likely than children in the control group to have a dental visit in the six months following their visit with WIC staff.

Awardee leaders were less certain about the influence of MCPP on the quality of dentists' service delivery or on collaboration between dental and medical providers. Although leaders reported success in disseminating educational materials to dentists, they did not have access to information about how dentists changed their service delivery in response. Two interviewees cited statewide increases in the number of children who had a dental visit or received an oral health service in the past year, but one-third were skeptical about MCPP's role in these increases because so few dentists were either engaged with the program or had registered to use MiDR.

When asked about their perceptions of the MCPP program's ability to achieve its three targeted outcomes, program leaders and staff interviewed in summer 2017 said they strongly believed the program had increased the number of children who had received preventive oral health care, primarily through PCP training and TA, WIC training, and SmileConnect.

Interviewees were hopeful that the MCPP had reduced the incidence of dental caries, and cited improvements in access to preventive services and the provision of oral health education to children and families as key drivers of any such reduction. Awardee leaders and staff were cautiously optimistic that taken together, increased access to preventive services and decreased numbers of dental caries will lead to lower spending on oral health care. Because the awardee had only recently received Medicaid claims data from the state, however, the awardee had not completed a cost analysis at the time of the virtual site visit in summer 2017.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

Awardee leaders, program staff, and providers who were trained thought the MCPP program positively impacted children's access to preventive oral health services, and thus may potentially have reduced the incidence of dental caries and overall costs for dental health care. However, dental providers' limited engagement with MCPP may limit the impact of the program on the incidence of dental caries, dental disease, and spending on dental care. For example, the program's inability to garner high levels of coordination between medical and dental providers, and the low rates of MiDR use—particularly among dental providers—likely hindered information sharing and referrals. This may have meant that only a few children got the benefits of consistent and long-term dental care through an established dental home. It will also be difficult to demonstrate improvements in oral health or decreases in its costs as a result of getting more preventive services in the three-year time frame of the cooperative agreement, because the impact of preventive care is more long term.

Even if an impact analysis of the MCPP program did demonstrate increases in children's receipt of preventive oral health care, decreases in the incidence of dental caries, and reductions in oral health care costs, two factors have implications for the interpretation of results. First, the multifaceted nature of the MCPP intervention would make it challenging to isolate the MCPP component, or a specific activity within a component, that most strongly influenced outcomes. Second, the continued expansion of Healthy Kids Dental—Michigan Medicaid's dental benefit program—during the time of Altarum's cooperative agreement would confound results of an impact study. As of summer 2016, 83 percent of Michigan dentists were participating in Healthy Kids Dental and were therefore willing to serve children under age 21 who were covered by Medicaid. The concurrent nature of the MCPP and Healthy Kids Dental would make it difficult to determine whether observed outcomes were associated with MCPP components or with children's access to dental care through newly acquired insurance coverage.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Altarum's MCPP program.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Altarum

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018	Not applicable; beneficiaries will be attributed to providers, and sample will be very large
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	2,932
Likelihood of all-cause hospitalizations	2,219
MDE sample size requirement to detect 20% effect	
Total expenditures	733
Likelihood of all-cause hospitalizations	555
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	Some issues, but probably surmountable. Expect to select a comparison group
Do claims identify the primary expected effects?	Some effects observed in claims data but important effects likely missing
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician and staff surveys
Implementation data that will be analyzed	Google analytics; some utilization data

We anticipate being able to construct a valid comparison group by drawing from the pool of practices that were not yet participating in the program during the first 18 months of program operations. Although comparison and treatment providers will come from different areas of the state, we expect to be able to draw a well-matched comparison group by matching on provider characteristics, including variables related to the provider's location (such as urbanicity) as well as characteristics of the patient panel. We will likely need to match at the provider level rather than at the practice level due to the difficulties involved in identifying practices in the claims data. We will then construct an analysis sample consisting of beneficiaries who have visited treatment or matched comparison providers.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

By the end of the third program year, Altarum had obtained Medicaid claims data and started using the data to develop a formal value model (that is, evaluate the potential benefits and costs of its program in order to justify the cost of the financial incentives proposed in the payment model). Altarum's proposed payment model is an enhanced FFS model with the potential for bonus payments. It features the following elements: (1) higher reimbursement rates for both preventive oral health services and dental treatment, and (2) a bonus payment to dental providers who reach a yet-to-be determined threshold (that is, a percentage) of their pediatric patients receiving dental preventive care.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: (1) the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor and (2) narratives and payment model reports the awardee submitted to the implementation and monitoring contractor. We did not interview the awardee about its payment model because it did not advance from the development stage to active negotiations with one or more payers.

B. Description of the payment model

The awardee is proposing an enhanced FFS payment model with the potential for bonus payments. The model has two main features: (1) increased reimbursement rates for both preventive oral health services and dental treatment services, and (2) a bonus payment to dental providers that reach a yet-to-be-determined target threshold (that is, a percentage) of children receiving dental preventive services. It is not clear what the awardee plans to use as the denominator population for calculating this percentage—whether it is all Medicaid-insured children seen by the dental practice or all Medicaid-insured children residing within a certain geographic area. The target population for the payment model is children who are under age 3 and enrolled in Medicaid, while the awardee's target population for the program is children up to age 17, with a focus on those age 5 and younger.

The services subject to increased fee-for-service reimbursement for primary care providers include oral health screenings and fluoride varnish applications. Dentists will receive higher fee-for-service reimbursement for dental exams and preventive oral health services at age 1. It is not clear what specific preventive oral health services will be eligible for the increased reimbursement, or how this will differ from the increased reimbursement already offered to dentists through the Healthy Kids Dental program.

C. Status of the payment model

At the end of the third year of the cooperative agreement, Altarum was continuing to conduct analyses to support the development of its payment model and was in conversation with Michigan Medicaid. To date, the awardee has obtained Medicaid claims data and has started using the data to evaluate the costs and potential benefits (in the form of improved oral health outcomes and associated savings) of its program in order to develop a formal value model to justify the cost of offering the financial incentives proposed in the payment model.

D. Factors associated with the development of the payment model

Although the use of MiDR is not a requirement of the payment model, the awardee believes it may advance the success of the payment model (for example, by increasing communication between providers, by providing better documentation of services provided, or as a source of data for evaluating impact). Additional information on the sustainability and scaling of MiDR is provided in Chapter V.

The awardee reported that the biggest challenge to developing a payment model has been in demonstrating the long-term benefit of promoting preventive oral health services for young children. The awardee also reported that even though its program works to increase referral rates to a dentist by age 1, many dentists are reluctant to accept young children into their offices as patients. Lastly, the awardee reported that dentists are unfamiliar with being paid outside of a fee-for-service model. Specifically, dentists typically do not monitor their population and measure performance, which would be a necessary step for dentists to determine if they are on track to receive the bonus payment for achieving the threshold level of preventive services.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Altarum made significant strides to promote the sustainability, scaling, and replication of the MCPP. While still pursuing its payment model to ensure the program's long-term sustainability, the awardee achieved three major milestones that will help sustain, scale, or replicate one or more of the program's components (that is, training and TA; public health improvement and health IT tool, such as MiDR; and participant and family engagement). First, Altarum secured state funding to continue MiDR. Second, state policy changes required broader adoption of MiDR by dental providers. Third, the awardee was working with UCLA to replicate and enhance the program's training and education and health IT components in LA County. However, Altarum had not had significant success in securing funding to sustain the training and TA functions in Michigan or replicate them beyond Los Angeles. The awardee also scaled the SmileConnect crowdsourcing website (which connects communities with dental supplies and services) to 30 states, and decided to brand the entire MCPP program as SmileConnect to signal the interconnectedness of all of the program components and the program's transferability to other states. The awardee planned to use its six-month, no-cost extension to continue working on its payment model, identify other funding to support the program components, and promote the program to other areas.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, Altarum had begun actively pursuing strategies for sustaining its program, but had not yet implemented these strategies. With regard to training and education, the model itself was designed to support sustainability of providers' behavior changes at the practice level. The awardee (1) was in the process of engaging philanthropic funders to support maintenance of the SmileConnect crowdsourcing website, (2) had plans to apply for other grants from the federal Health Resources and Services Administration or other sources to support the training and TA component, and (3) was analyzing program data to develop payment models for the core program and to foster its replication in other states (see Chapter IV). The awardee had not scaled or replicated the program as a whole, but by the end of the second program year, it had secured 16 dental and dental hygiene schools across eight states to participate in SmileConnect.

C. Implementing the SSR plan: progress and changes

While pursuing its payment model (higher reimbursement rates and bonus payments, as reported in Chapter IV) over the past year, the awardee pursued other activities to foster SSR to varying degrees for the program's three main components: (1) training and TA to medical staff, (2) health IT (a dental registry, MiDR), and (3) participant and family engagement (through outreach to schools, early education programs, and public health programs such as WIC). It had three major SSR achievements, one each in the areas of sustaining, scaling, and replication.

Sustainability. First, the awardee, along with Delta Dental, successfully lobbied the state legislature to provide $500,000 annually for ongoing monitoring and maintenance of MiDR. This reportedly was a trying and time-intensive effort, in part because many legislators wanted more evidence that the investment would save significant amounts of money. Although the awardee was unable to produce evidence of such savings, it was able to convince enough legislators that it was the right thing to do and prudent to continue with MiDR (that is, it would be a waste to end what was already developed).

Respondents expected that MiDR monitoring, maintenance, and promotion would be maintained through combined efforts from some state Medicaid and Altarum staff, but respondents did not yet know whether the state or Altarum would take ownership of MiDR. Also, Delta Dental was leading a large team working on the future of MiDR and determining the extent to which the registry would remain in its current form or change, potentially to capture additional data elements or populations (see paragraphs that follow the Scaling lead-in below).

However, with fewer available resources than the program had during the HCIA R2 award, respondents expected a decline in activities promoting MiDR and educating and training providers on it. For example, Altarum ended its training for PCPs in May 2017 and has already experienced a "big drop" in the number of new Smiles for Life certifications. Delta Dental decided to continue educating dental providers and public health programs, including WIC, about MiDR—but with only one staff member compared to the eight or nine staff Altarum had through the grant. Altarum and Delta Dental were considering ways to provide training and TA on MiDR to dentists because more dentists were expected to use the registry (see Scaling), but any plans would have to overcome a lack of funding for training. Further, Altarum has not gotten enough funding to sustain the portion of the MCPP that offers training and TA to medical

providers, with the exception of some scaling and replicating activities in LA County (see paragraphs following Replication below).

Delta Dental planned to sustain its work on the participant and family engagement component with WIC generally (for example, through the WIC BRUSH program) for about one more year.

Scalability. The second major development toward SSR was that the state of Michigan mandated that dental insurers—via Medicaid contracting arrangements made as of January 1, 2018— require dentists in their networks to adopt MiDR. Respondents thought this was a useful way to embed the MiDR functions into the vendor contracts and ensure more dentists use it going forward. However, one respondent expected the requirement would mean fewer dentists would be willing to treat Medicaid patients because Medicaid payment rates are already low, and adding this requirement would increase the costs of practicing dentistry. The awardee and Delta Dental were considering ways to expand MiDR to include commercially insured children (so that it would be open to all children), which would also help encourage more medical and dental providers to use MiDR.

Also, some entities participating in the MCPP were interested in scaling some program components. The awardee was attempting to build a real-time integration between MiDR and Management Information for WIC to end the need to upload dentist information manually; this integration would enable rapid progress and expansion of MiDR. The awardee obtained funding from the Delta Dental Foundation and the Dittrick Medical Center Foundation to support this effort, although the level of actual integration that was achieved depended on the funding level. Further, the Michigan WIC program wanted to implement the WIC BRUSH at all of its sites. The awardee was working with Delta Dental and McMillen Health (a health education organization) to provide BRUSH training to WIC staff.

SmileConnect was on track to scale nationally since its launch in Michigan in February 2016. Between March and August 2016, SmileConnect grew from use by dental and dental hygiene schools in Michigan to include service providers from eight states. During that time, Altarum reached out across the nation to philanthropies and companies for donations. By the end of the third program year, Altarum reported that SmileConnect operated in 30 states, partnered with 34 dental and dental hygiene schools, and affected an estimated 31,000 children by connecting communities in need with supplies and services.

Replicability. The third major development toward SSR is that Altarum was working with UCLA to replicate the MCPP program in LA County. According to the Altarum leaders we spoke with in interviews, it appeared that LA County would replicate the same "multi-level, comprehensive" program approach (including the training and TA, dental registry, and participant and family engagement) but that Los Angeles would adjust the program to its own circumstances and barriers and enhance it—for example, by providing program materials in Spanish and giving dentists more education on it.

Altarum planned to retain a central role in facilitating the LA County replication. For example, it expected to lead the training component, but potentially streamline and improve it by switching to a "flipped classroom approach," in which trainees would review more formal

training materials before the in-person training sessions, and then use the training sessions to talk through the specifics of implementation, such as determining staff roles and customizing workflows. Altarum was also helping develop a dental registry similar to MiDR, called the Los Angeles Dental Registry, or LADR. The awardee viewed such enhanced replication as a next step in the program's evolution and hoped that its involvement in replication would build momentum and support for sustaining the program.

Indeed, the awardee thought that efforts to replicate the program were just beginning, and it hoped to see replication on a national scale. In fact, the awardee was in discussions particularly about the training and TA component with Texas providers and Delta Dental in Washington State, and was thinking of partnering with the American Academy of Pediatrics (with whom it has a personal connection) to extend their reach further than it could by working state by state. Also, Delta Dental recently won the Medicaid contract in Arkansas, which could foster replication in that state.

To help foster replication, the awardee rebranded the entire program (education, training/TA; dental registry; and participant and family engagement) under the umbrella term Smile Connect and trademarked that name so the program would not seem limited to Michigan. Although the awardee did not want to be prescriptive, Altarum planned to encourage others to implement the program as a whole, instead of just using certain pieces, because the different components support each other and amplify the effect of each.

The awardee planned to use its no-cost extension period to establish its payment model and increase awareness of the program's success across markets nationally. The awardee expected to evaluate its claims and other data to "develop our story so that we can really share it and define what can be scalable to other regions and other states."

D. Factors associated with progress toward implementing the SSR plan

In addition to the factors affecting the awardee's ability to implement its payment model (as discussed in Chapter IV), interviewees also discussed several other facilitators and barriers to SSR. The awardee noted that Michigan's policy environment supported the program by (1) providing a high level of advanced technology and a data warehouse, (2) implementing policies that encourage dentists to treat children covered by Medicaid, and (3) paying physicians to screen children for dental problems and apply fluoride varnish (although some states pay more than Michigan does). Even with this support, the awardee faced significant challenges in getting the state onboard to help sustain the program and in needing to educate payers that the program services were reimbursable. On the other hand, the CME and MOC credit requirements for state licensing are the same all over the country, so the awardee saw no ongoing education or licensing barriers to expanding the PCP training component nationally, and thought the promise of these credits would give providers an incentive to participate. The awardee told stakeholders in other states who were interested in replicating the program about these features, so they would understand how the broader state policy context could either facilitate or hinder their efforts to replicate the program.

In addition, because Michigan strictly regulates dental insurance, Delta Dental had long had an active government relations office with strong connections to legislators and state agencies.

The ability to leverage these connections and relationships helped secure the ongoing funding for MiDR. However, the awardee was unable to gain Michigan DHHS's support in advocating for this funding, which may have slowed the approval process.

Respondents expected some challenges in maintaining MiDR. The lack of clear ownership of MiDR by a single entity was a barrier to sustainability and scalability. Respondents held mixed views as to whether $500,000 was the right amount to sustain and scale MiDR. The awardee was concerned about having the capacity to register all dentists in a short time period and to provide the necessary support to help them use MiDR.

Reportedly, a key reason behind Michigan's decision to expand the bidding for Medicaid dental contracts to additional plans was the new CMS requirement that Medicaid dental programs integrate patients' medical and dental information. Delta Dental lacks a medical insurance component, and expected other insurers that offer both medical and dental services to have a competitive advantage for the new contracts. If many enrollees went to other insurers, it could take time for other players to get up to speed on the program, possibly disrupting the Healthy Kids Dental program (that is, altering its form), and causing SSR plans to stall, particularly if the new plans had trouble building their dental networks and needed considerable resources to meet reporting requirements.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Altarum's MCPP program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to beneficiaries from the pool of practices that were not yet participating in the program during the first 18 months of program operations whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Amerigroup

April 18, 2018

Grace Anglin (Mathematica Policy Research)
Michaella Morzuch (Mathematica Policy Research)
Lisa Lines (RTI International)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Amerigroup, the sole Medicaid managed care provider for Georgia's foster care program, received an HCIA R2 award to implement the Coaching and Comprehensive Health Supports (COACHES) program (Table I.1). Through COACHES, Amerigroup and its partner, Families

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Amerigroup received a six-month no-cost extension that ended February 28, 2018.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

First, provided intensive coaching services for youth who lived in participating counties in Georgia and were about to transition out of foster care. Youth in foster care often experience significant mental and emotional stress during childhood and lack a continuous adult support system. As a result, these youth are often ill-equipped for the transition to adulthood. The COACHES program paired participating youth with a coach employed and trained by Families First to help prepare them for this transition by educating them about the health care and social services systems, helping them build life skills (such as renting an apartment and household budgeting), and supporting them as they advocated for their own needs. The COACHES program was youth-directed—an innovative approach to working with foster care youth. Unlike many other programs for this population, Families First made participation in COACHES voluntary and youth determined the focus of their work in the program.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The COACHES program connected youth who were about to transition out of foster care with an employed, trained coach who taught them how to access, coordinate, and manage health and social services on their own.
Major innovation	Unlike many other programs for foster care youth, Families First made participation in COACHES voluntary and participant-driven. For example, participants determined how often they met with their coaches and the focus of their work in the program.
Program components	• Patient and family engagement • Care management services • Outpatient care coordination
Target population	Youth ages 17 to 20 who have been in foster care for 12 months or longer,[a] have a documented history of behavioral health needs, and reside in participating counties.[b] Youth who resided with foster families or in group homes had to get permission from their foster parents or group home staff to participate in the program.
Theory of change/ theory of action	Amerigroup hypothesized that youth who worked closely with a coach would better understand what services they needed and how to access them. Participants would then increase use of primary care, pregnancy prevention services, and educational and employment programs, which in turn would result in better health and social outcomes, as well as lower health and social service costs.
Payment model	Amerigroup's partner, Families First, has secured a $2 million contract from the Georgia Department of Social Services to sustain the intervention beyond the cooperative agreement. Amerigroup is not a party to that contract and does not have any plans to develop a payment model for work related to the program.
Award amount	$5,833,492
Effective launch date	March 1, 2015
Program setting	Community- and home-based
Market area	Urban, rural, and suburban
Market location	Participating counties in Georgia[b] (Baldwin, Bartow, Bibb, Carroll, Cherokee, Clayton, Cobb, Coweta, Dawson, DeKalb, Douglas, Fayette, Forsyth, Fulton, Gwinnett, Hall, Hancock, Harris, Henry, Houston, Jones, Macon, Newton, Muscogee, Paulding, Peach, Pike, Polk, Randolph, Rockdale, Spalding, Steward, Talbot, and Taylor)

Table I.1 *(continued)*

Program characteristic	Description
Target outcomes	• Improved health literacy and ability to navigate the health care system • Increased use of primary care and preventive services • Higher educational attainment and increased employment • Improved connections to peer and adult social supports • Improved life skills (including renting an apartment and household budgeting) • Increased knowledge of legal and juvenile justice systems • Decreased health and social service costs

[a]Amerigroup expanded from only serving youth living in group homes to include youth enrolled in independent living programs early in Year 1. Later in Year 1, the program also began to include youth who resided with foster families.

[b]Amerigroup expanded from 6 to 11 counties in Year 1, from 11 to 18 in Year 2, and from 18 to 34 in Year 3.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we concluded that the awardee was partly successful in implementing its program by the end of the initial three-year cooperative agreement. Amerigroup was successful at meeting its enrollment targets, and it hired and retained the number of coaches it needed to implement the program. Those coaches generally followed implementation protocols. However, the median length of time youth stayed in the COACHES program was three months, considerably less than the 12 to 18 months the awardee expected. Coaches also struggled to engage caseworkers at the Division of Family and Children (DFCS) and foster care providers (such as staff at independent living programs or group homes).

Impact evaluation. Given that Amerigroup anticipates primarily affecting outcomes that are not observable in the Medicaid claims data (such as the participant's health knowledge, education, employment, and life skills), we do not anticipate being able to conduct a rigorous impact analysis for Amerigroup.

Payment model. Amerigroup did not and does not plan to develop a payment model because its partner, Families First, secured state funding to sustain the program.

Sustainability plans. Families First secured a $2 million contract from the Georgia Department of Social Services to sustain the COACHES program beyond the cooperative agreement. Amerigroup is not a party to that contract. Families First expected to maintain the program largely "as is" and to scale it to new populations, including younger adolescents and Department of Juvenile Justice (DJJ) clients. Families First trademarked the program name and some of its components in an effort to promote its replication while maintaining its integrity.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of staff on their perceptions of the program's effect on the delivery of care. The survey was fielded from July to October of 2016 with all 31 staff members who were part of the program at the time, and achieved a response rate of 83 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partly successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

Amerigroup met its enrollment targets, enrolling 860 direct participants, or about 119 percent of its final three-year projections. Moreover, the awardee was able to hire, train, and retain the number of coaches it needed to implement the program, and those coaches generally followed implementation protocols. However, participants only spent a short time in the program. Their median length of time in the COACHES program was three months, considerably less than the 12 to 18 months the awardee had expected. Additionally, coaches struggled to engage community providers, which limited their ability to coordinate care across the various physical and mental health and social service systems that participants encountered. Amerigroup received a six month no-cost extension—which ended in February 2018—to develop a customized database to more accurately and efficiently document, track, and report on program services and to pilot scaling the program to DJJ clients.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

Amerigroup and Families First originally planned to get referrals from DFCS caseworkers and foster care providers (such as staff at independent living programs or group homes). After receiving fewer referrals than expected early in the program, Amerigroup and Families First adjusted their recruitment strategies to focus on the youth themselves instead of only seeking third-party referrals. In particular, during Year 1 Amerigroup started directly promoting the program to youth, with Families First advertising the COACHES program to potential participants at events and meetings attended by youth in foster care.

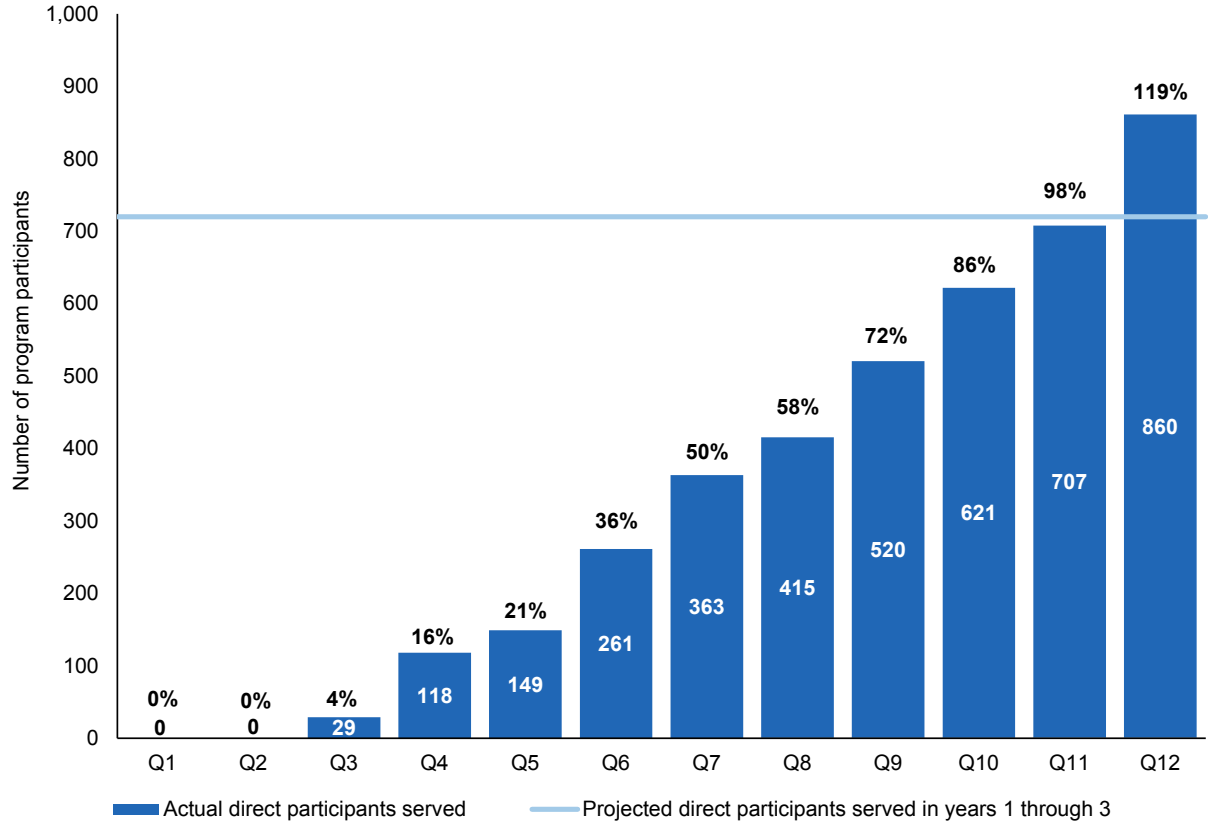
The awardee also expanded the eligibility criteria to include youth who lived in other counties and/or did not live in group homes. Amerigroup expanded from 6 to 11 counties in Year 1, from 11 to 18 in Year 2, and from 18 to 34 in Year 3. Amerigroup expanded from only serving youth who lived in group homes to include youth enrolled in independent living programs early in Year 1, and to include youth who lived with foster families late in Year 1.

After youth were referred to the program, the program manager at Amerigroup verified their eligibility. A coach next met in person with youth who were deemed eligible in order to describe the program and its benefits and to encourage them to participate. Youth who wanted to participate signed a consent form. Participants living in group homes or with foster families also had to receive consent to participate from group home staff or their foster parents.

b. Evidence of enrollment effectiveness

Amerigroup exceeded its enrollment goals, which did not change over the course of the cooperative agreement. After experiencing low enrollment rates in the first few months of the program, the COACHES program steadily increased the pace of enrollment. Overall, the awardee reported that it enrolled 860 direct participants from March 2015 (when it launched its program) through August 2017, which represents about 119 percent of its final three-year projections (Figure II.1).

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. In reports to the implementation and monitoring contractor, Amerigroup classifies their participants as indirect. Since these participants' receive HCIA R2-funded services, we reclassified Amerigroup's participants as direct participants in this report. Amerigroup's enrollment target did not change over the course of the cooperative agreement.

Although Amerigroup met its enrollment goals, it may not have enrolled the youth in foster care who were most in need of its services. The program relied heavily on self-referrals to meet its enrollment targets, and some coaches reported that self-referring youth were relatively high functioning compared to other youth in foster care. Amerigroup also did not rely on medical records or behavioral health screening to determine if youth met the eligibility criteria that required participants to have a documented history of behavioral health needs. COACHES staff acknowledged that the program did not focus exclusively on youth with the greatest needs, but also noted that all youth in foster care generally do have higher needs than other youth.

c. Barriers and facilitators associated with enrollment effectiveness

Amerigroup and Families First said their shift to focus on self-referrals helped the program surpass its enrollment target. Staff attributed the high rate of self-referrals to the youth-driven

nature of the COACHES program and the skill and enthusiasm of program staff. COACHES staff and other DFCS stakeholders believed that because the program was voluntary and youth-directed, it motivated youth in foster care—who often live highly scheduled and managed lives—to actively engage in COACHES. Moreover, program leaders reported that the coaches' enthusiastic and youth-oriented approach to recruitment encouraged participation. For example, coaches incorporated games—such as basketball and obstacle courses—into recruitment events.

A barrier to enrollment throughout the program was that foster care providers referred far fewer participants to COACHES than Amerigroup initially expected. In addition, some group home staff or foster parents were hesitant to approve participation for youth who had self-referred to the program. Program staff attributed the low interest levels to providers' limited time and resources, their lack of knowledge about the program (a continuous problem because of the turnover in foster care providers), or their perception that the program duplicated services they were already offering.

> "We've built enough of a brand and our work has spoken for itself [enough] that [relationships with group home providers] have definitely improved."
>
> —Program staff

To make foster care providers more comfortable with the program and help overcome this barrier, Families First centralized its outreach approach during the second program year. Only program leaders (as opposed to any staff) educated foster care providers about the program, and one staff person served as the main point of contact for foster care providers. Program leaders noted that face-to-face meetings with local staff and sharing information on youth's participation in the program with foster care providers were critical to building relationships. In addition, Families First instituted an electronic referral process (in which DFCS providers could submit referral forms by email) to make referrals easier and more secure. As a result of these efforts, the awardee reported a steady increase in the number of referrals to the program over the second and third program years.

2. Delivery of program services

a. Description of and changes to service delivery model

COACHES was a new service delivery model for both Amerigroup and its partner, Families First. Amerigroup was responsible for overseeing the program, verifying participants' eligibility, and performing analytics on Medicaid claims data. Families First was responsible for hiring and training full-time coaches who worked directly with participants. The awardee developed a series of implementation protocols that described in detail the steps coaches should take when working with youth, refining those protocols as needed over the course of the program to provide more guidance to staff. This section briefly describes how the awardee envisioned COACHES implementation and highlights changes made to the service delivery model.

After youth were enrolled in the program, their assigned coach was supposed to contact them within 24 hours to schedule an in-person meeting. During that following meeting, coaches were charged with completing a series of standardized psychosocial and trauma assessments to understand the youth's strengths and needs. The coach also asked youth to set at least one personal goal in each of the program's five focus areas: (1) education, (2) employment, (3) mental health, (4) physical health, and (5) pregnancy prevention. Then, the youth and coach

worked together to develop a coaching skills plan that laid out the steps that the youth could take to meet his or her highest priority goals. Amerigroup intended all youth to have a preliminary coaching skills plan developed after their first meeting with their coach and a refined plan in place after the third meeting.

Amerigroup and Families First anticipated serving youth in the COACHES program for 12 to 18 months. For the duration of enrollment, coaches were expected to meet with participants to review and update their plan and help them make progress toward their goals. Coaches were supposed to proactively reach out to participants at least once a week, though participants ultimately drove the frequency, mode, and focus of those meetings. Amerigroup and Families First envisioned coaches using the following strategies to help participants meet their goals:

- Educate participants about the health care and social services systems and help them build life skills (such as how to rent an apartment and how to construct and live on a household budget). Coaches were trained in motivational interviewing and the transition-to-independence model, which included helping youth to break large goals down into smaller milestones and empowering them to take concrete steps toward achieving the milestones. [4]
- Refer participants to community-based services and supports (such as their Amerigroup health care coordinator, employment and education services, or weight loss programs). Before a participant could be referred to a community-based program, Amerigroup required the program to be vetted by leaders at the COACHES program.
- Facilitate communication and information sharing between youth, their different service providers (such as medical, behavioral health, and foster care providers), and informal supports (such as religious leaders or family members). To help coordinate care, the awardee planned for coaches to share coaching skills plans with these providers and attend DFCS-established family team meetings.

The awardee made several refinements to the program over time.

- In the second program year, coaches started hosting group education sessions (known as huddles) with participants. These were held about every three months and supplemented individual coaching. The huddles were in-person meetings focused on areas of particular interest or need for participants, including employment opportunities and contraception.
- In the final program year, to account for a shorter-than-expected enrollment period, the awardee reduced the number of initial assessments and the frequency of all assessments. Coaches said the change allowed them to complete care plans sooner and, by reducing the paperwork required to participate, helped keep youth engaged.
- Also in the final program year, the awardee added the option for youth in foster care who disenrolled from the full COACHES program to continue periodic check-ins with their coach. These contacts were youth-initiated, less intensive, and less frequent than contacts for full enrollees.

[4] For more information on the Transition to Independence Process Model, visit http://www.tipstars.org/.

b. Evidence of service delivery effectiveness

Staffing and training. The awardee was effective in hiring and training COACHES staff. Amerigroup and Families First hired enough staff to launch the COACHES program on schedule. As enrollment grew, program leaders hired and retained enough staff to maintain youth-to-coach ratios around their ideal of 15 to 1. Moreover, in response to the staff survey fielded from July to October 2016, 84 percent of staff reported that their workload was about right, and staff reported being very satisfied (46 percent) or somewhat satisfied (42 percent) with their role in the COACHES program.

Before they started working with the youth, coaches received about 100 hours of in-person training on evidence-based strategies, including techniques for motivational interviewing and tools based on the Transition to Independence Process Model. When asked about training on the 2016 staff survey, 95 percent of staff agreed that they learned new skills during training, and 96 percent agreed that training helped improve their job performance. Specifically, during annual qualitative interviews, staff indicated that COACHES trainings gave them the skills they needed to encourage participants to take control of their service needs without explicitly directing them in how to do so.

Delivery of intervention services. The awardee was generally successful at delivering intervention services, although implementation varied somewhat depending on the coach. During interviews, coaches said they used the program's implementation protocols to guide the delivery of assessments and help them know how and when to reach out to youth.

Staff also consistently reported using the evidence-based practices integrated into the COACHES model—most notably motivational interviewing—to encourage participants to manage their health and social services. Moreover, awardee leaders—drawing on data from focus groups they conducted with participants—reported that most coaches followed established implementation protocols, adhering to the model in terms of the frequency and type of interactions between coaches and youth, and maintaining the youth-driven nature of the program.

Awardee leaders also noted, however, that in the first two program years, coaches often took more than three meetings with youth to complete all the assessments and develop an initial coaching skills plan. Coaches reported that the awardee's decision to reduce the number of assessments in the third program year helped them complete those plans sooner. The awardee leaders also reported some variation in delivery of intervention services across coaches, with some focus group participants rating their coaching skills plans and meetings higher than other participants rated theirs.

Engagement of program participants. The awardee did not successfully engage participants for the length of time it planned to. The median length of time that coaches worked with youth was three months, considerably less than the 12 to 18 months the awardee had planned on. Moreover, the awardee reported that coaches devoted significant time to scheduling meetings as opposed to delivering services. On the 2016 staff survey, coaches were asked to indicate whether items on a list of given factors were a major barrier, a minor barrier, or not a barrier to COACHES achieving its goals. Over half of staff (58 percent) reported that resistance to the program by participants was a barrier to the program, though most said it was a minor

barrier (50 percent). Program leaders reported that the program was more successful engaging youth in rural areas because there are fewer supports for foster care youth in rural communities.

Although coaches faced challenges getting youth to attend meetings, they generally reported that—once at individual or group learning sessions—participants actively engaged with their coaches. According to interviews with staff and data reported to the implementation and monitoring contractor by the awardee, youth focused most often on education- and employment-related goals in their work with their coaches. Although they were a less intensive focus, program staff also indicated that youth set health- related goals, such as identifying a primary care provider or discussing prescribed psychotropic medication with their provider.

Recruitment and engagement of providers. To help youth coordinate the services they received from various medical and behavioral health providers, child service agencies, and community organizations, coaches tried to meet and work with these organizations. However, coaches reported that limited engagement on the part of these organizations made it challenging to help youth coordinate services and achieve their goals. Although some staff reported that engagement got better over time, coaches generally reported they had few opportunities to discuss care plans with other providers over the course of the award. In some cases, providers even undercut the coaches' attempts to work with youth. For example, one group home restricted the times that coaches could meet with participants at home. In the 2016 staff survey, lack of buy-in or difficulties coordinating with community providers were the most commonly reported barriers to the COACHES program achieving its goals. Most notably, 79 percent of staff reported that lack of buy-in from legal guardians, state agencies, or referral sources was a major barrier (29 percent) or minor barrier (50 percent) to COACHES' achievement of its goals.

> "When the caseworkers don't call us back or respond to our emails, they make it hard for us to go to the next step. . . . If we need something from the caseworker to help [youth] with their goals and we're not able to reach them, it makes the process difficult."
>
> —Program coach

c. Barriers and facilitators associated with service delivery effectiveness

The awardee's leaders and staff described two primary facilitators to implementing the service delivery model: (1) training and support for coaches and (2) a focus on continuous quality improvement.

Each program year, coaches reported that ongoing training opportunities and a supportive supervisory structure contributed to their job satisfaction and their ability to effectively engage with youth. Coaches said they felt comfortable raising questions and problem-solving with their supervisors, and that Families First's supportive culture helped them cope with the emotional strain of working with a high-need population.

> "The management team hears concerns we raise, and they address them to make our job even easier than it already was. So, all we have to do is focus on the youth and the things they are going through."
>
> —Program coach

Coaches also said their colleagues, including those not directly involved with the program, helped them think through challenging situations and identify community resources for participants. On the 2016 staff survey, most respondents strongly agreed (67 percent) or somewhat agreed (25 percent) that they felt supported by their colleagues and management in

doing their job. Moreover, respondents strongly agreed (83 percent) or somewhat agreed (17 percent) that Families First and Amerigroup gave them the equipment, supplies, and resources they needed to do their job. During interviews, coaches said their implementation was supported by the in-house resources at Families First, such as subject-matter and information technology experts. For example, an education consultant at Families First helped one youth identify and secure transportation so he could stay at his school after moving to a group home in a different district.

> "It is all about tightening up our processes—making sure everything is documented, making sure that we have a truly canned product that can be modeled elsewhere. . . . Just ensuring that all of our processes and ducks are in a row."
>
> —Program leader

Amerigroup and Families First's focus on continuous quality improvement also facilitated implementation of the service delivery model. In the second and third program years, the awardee focused on standardizing its service delivery structure. For example, program leaders continuously updated the COACHES program manual to refine procedures and developed program fidelity checklists for coaches. Families First also hired an evaluation specialist in the third program year charged with helping improve the program's assessment and self-monitoring processes. At the end of the third program year and into it's no cost extension period, the awardee also started to develop a customized database to more accurately and efficiently document, track, and report on program services, which staff consistently indicated was difficult to do in the program's electronic record system.

Leaders and staff at the COACHES program also reported two major barriers to implementing the program: (1) keeping participants engaged and (2) coordinating services with DFCS case workers and foster care providers.

In terms of participant engagement, coaches indicated in each program year that participants' hectic schedules or behavioral health challenges sometimes meant that youth did not respond to their coach's phone calls or missed their appointments. In addition, coaches reported that youth often decided to leave foster care at age 18, making them ineligible for the program. Coaches indicated that convincing youth to remain in care was difficult because many youth have a deep distrust of the foster care system.

> "Reducing the number of assessments actually helped engage youth because I didn't have to go through all these surveys that they've done a million times. They got to get right down to what they wanted to work on instead of spending three or four sessions getting through all the surveys before getting started with the actual coaching."
>
> —Program leader

Staff refined the program to improve youth engagement and to account for the shorter-than-anticipated enrollment period. For instance, coaches streamlined the assessment process and focused on fewer, higher-priority goals with youth to maximize the program's impact while they were enrolled. Coaches also tested youth-friendly communication and education strategies. For example, to ease anxiety about the program, coaches encouraged youth who set challenging goals to take small, manageable steps toward achieving them, such as developing a resume or reviewing a medication list. As another example, coaches started texting youth instead of calling them to schedule in-person meetings and to check in on progress toward their goals. In addition, coaches realized that youth appreciate opportunities for peer-to-peer interactions, so

they began to host group education sessions. Coaches reported during interviews in the third program year that these strategies were increasing participant engagement.

As for challenges coordinating care, program staff reported similar reasons for low engagement of foster care providers' in care coordination as they did for low referral rates—most notably that foster care providers had limited time and resources and frequently changed positions. The strategies intended to increase referrals, outlined in Section II.B.c, were also intended to encourage providers to remain involved in the program. In addition to these strategies, Families First instituted new procedures in the second program year requiring coaches to reach out more proactively to providers to discuss the youth's coaching skills plans. Moreover, Families First also offered foster care providers free training opportunities to motivate them to work with coaches. Through these combined efforts, some coaches reported better coordination with foster care providers during the second and third program years.

For more details on the factors influencing COACHES implementation, please see the awardee-specific narratives in the first and second annual evaluation reports.[5]

C. Assessment of perceived program effects on the delivery of care and outcomes

Program staff thought COACHES positively impacted intermediate outcomes for youth, such as encouraging them to stay in foster care, teaching them about the health care system, and helping them get ready to apply for jobs. In the 2016 staff survey—fielded around the start of Year 3—most staff reported that the COACHES program was very effective (71 percent) or somewhat effective (17 percent) at achieving its goals. Similarly, nearly all staff strongly agreed (71 percent) or somewhat agreed (25 percent) that the COACHES program was making a difference in meeting critical needs in the community. Specifically, 100 percent of staff reported that the program had a positive impact on participants' ability to achieve goals in three primary focus areas for the program: employment, education, and health. During qualitative interviews, program leaders and coaches also said the program had enhanced participants' decision making capability and empowered them to manage their own health and social service needs. Examples of this follow:

> "When I asked the participants about how they felt they may have benefited from the [COACHES] program, it really had little to do in many cases with medical care or having a medical care home and really so much more about making it through early adulthood and late adolescence without a parent."
>
> —Program leader

- A coach developed telephone scripts for a particularly shy youth that had difficulty interacting with adults to use when the youth called to set up doctor appointments or inquire about jobs. The coach used the telephone scripts to help guide her through those

[5] See: Anglin, Grace, Michaella Morzuch, and Amy Helburn. "HCIA Round 2 Evaluation: Amerigroup. Individual Awardee Second Annual Program Narrative." Prepared for the Centers for Medicare & Medicaid Services. Princeton, NJ: Mathematica Policy Research, January 2017. Also see: Anglin, Grace, Michaella Morzuch, and Amy Helburn. "HCIA Round 2 Evaluation: Amerigroup. Individual Awardee First Annual Program Narrative." Prepared for the Centers for Medicare & Medicaid Services. Princeton, NJ: Mathematica Policy Research, April 2016.

conversations. After several months in the program, the youth was able to make those calls without the coach's support.

- A coach helped a youth struggling with his self-taught GED program to identify why he was struggling and to think through solutions for furthering his education. The youth determined that a group learning environment—such as night classes—would work better for him, and he asked his DFCS team to transition him to a group class.
- A coach helped a youth who was considering leaving his group home after an altercation with a peer to think through strategies to diffuse the situation, and role-played those strategies with him. The youth settled the dispute and stayed in his placement.

Most often, program staff attributed the impacts they credited to COACHES to the youth-focused, strengths-based nature of the program. Staff believed that placing the nexus of responsibility for setting program goals on youth empowered them to make changes. Staff also reported that staying positive and avoiding judgmental statements allowed coaches to gain participants' trust and help them develop plans to meet their goals. Several coaches also reported that referring youth to community resources contributed to better outcomes.

During interviews, program staff expressed their strong belief that these intermediate outcomes would ultimately result in long-term impacts, such as higher rates of employment, lower DFCS costs associated with youth transitioning between placements, and reduced physical and behavioral health care costs. One respondent described the program as a series of small wins (such as learning to develop a resume and call an employer) that build up to larger gains (such as getting a job). Moreover, awardee leaders believed early results from their internal analyses pointed to more use of contraceptives and less psychotropic drug use, fewer emergency department visits, and fewer inpatient behavioral health stays.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Program staff thought the COACHES program had a positive impact on participants. However, the impact of the program on health-related outcomes may have been limited by the following factors:

- The awardee planned on a 12- to 18-month enrollment period, but the median length of time that participants were in the program was 3 months. Consequently, participants may have gotten less out of the COACHES program than they would have if they had stayed in it as long as they were expected to.
- Youth focused most on education- and employment-related goals, not health-related goals.

Even if the COACHES program (as designed and implemented) resulted in better health outcomes or lower health care costs, it will be challenging to demonstrate those outcomes through a rigorous impact analysis of Medicaid claims data for the following reasons:

- Staff stressed that COACHES most directly impacted intermediate outcomes—such as improving health literacy. Although staff believed these changes would ultimately bring

about reductions in health care costs, those impacts may not be observed within one year of the program ending (the period available for detecting impacts for the HCIA R2 evaluation).

- A large proportion of the participants self-referred to the program, and some coaches thought the program enrolled relatively high-functioning youth as a result. These youth may have had fewer needs and lower spending levels at baseline compared with youth who were referred by foster care providers, meaning there was less opportunity for improvement as a result of the program.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Amerigroup's COACHES program.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Amerigroup

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Projected final sample size	526[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	2,932
Likelihood of all-cause hospitalizations	2,454
MDE sample size requirement to detect 20% effect	
Total expenditures	733
Likelihood of all-cause hospitalizations	614
Participation/Selection bias of concern	Possible concern because participants may be higher functioning than average foster children
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	Too early to determine due to delay in Medicaid data
Do claims identify the primary expected effects?	Some effects observed in claims data but important effects likely missing
Core outcomes estimation method	Possible difference in differences on some outcomes
Primary reason for no rigorous evaluation	Too few beneficiaries to detect core outcomes, but may have be able to assess other outcomes
Survey data for treatment group that will be analyzed	Staff survey
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

At this point, we are not sure if we will be able to conduct a rigorous impact analysis for Amerigroup. It would be difficult for Amerigroup's program to affect claims-based outcomes during the model testing period because many of the outcomes that the awardee anticipates affecting are not observable in the Medicaid claims data (such as the participant's health knowledge, education, employment, and life skills). In addition, it is likely that the projected final sample size in Table III.1 (which reflects the number of treatment group members as of August 2017 when enrollment ended) is not large enough to detect effects on most claims-based outcomes. However, we might have a large enough analysis sample to be able to detect whether the program affected the participants' likelihood of receiving primary care and possibly whether they had any visits to the ED.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Amerigroup did not and does not plan to develop a payment model because its partner, Families First, secured state funding to sustain the program.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Although the awardee made some preliminary efforts to develop a value-based payment model for Medicaid that was tied to performance on designated process and outcome measures, it did not develop a fully specified model, and does not plan to continue to pursue payment models for the program.

C. Status of the payment model

The awardee's partner, Families First, secured a $2 million contract from the state of Georgia to sustain program activities. Amerigroup is not a party to that contract and has no plans to develop any payment models for ongoing work related to the program.

D. Factors associated with the development of the payment model

Amerigroup used existing value-based purchasing contracts—developed in its capacity as the sole Medicaid managed care organization for foster care youth in Georgia—to develop its preliminary payment model.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Families First secured a $2 million contract from the Georgia Department of Social Services to sustain the COACHES program beyond the cooperative agreement. Amerigroup is not a party to that contract. Families First expected to maintain the program largely as is and to scale it to new populations, including younger adolescents and DJJ clients. Families First trademarked the program name and some of its components in an effort to promote its replication while maintaining its integrity.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2 of the award, Amerigroup had minimal plans for sustaining its COACHES program. The awardee was planning to meet with potential funders, including state child service agencies the awardee had a relationship with. The awardee also refined its evaluation plan to help ensure it would have data on program impacts to support its conversations with potential funders. At that time, Amerigroup had not scaled its program, nor had it worked with other entities to help replicate the program.

C. Implementing the SSR plan: progress and changes

Sustainability. As discussed in Chapter IV, in the third program year the state of Georgia entered into a $2 million contract with Families First to sustain the COACHES program. As of September 2017, the state gives Families First $500 per COACHES participant per month. To receive funding for a participant, coaches must document four contacts with the participant per month, two of which must be in person. COACHES is expected to serve a cumulative number of 500 new clients under the state contract.

Amerigroup did not envision that the state would choose to contract directly with Families First to manage the program, but the awardee was confident that the program would remain successful under this new structure. Program leaders at Amerigroup and Families First did not expect the program design or approach to change in the future, with the exception of some slight modifications to serve a broader population and to align with state regulations (see below).

Amerigroup expected to use its six-month no-cost extension on activities that would indirectly support sustaining the program. This included building a new database to document, track, and report on program services—information that could be useful in garnering additional funding.

Scalability. The awardee planned to scale the COACHES program in two ways. First, starting during the no-cost extension, the awardee was piloting scaling the program to DJJ clients as a way to improve health and social service outcomes and decrease rates of recidivism. The awardee was testing the program for Amerigroup members age 17–20 who were in the sole custody of the DJJ (and were not eligible for the original COACHES program), lived in a community setting, and were willing to participate. Amerigroup indicated that this expansion would require training the coaches on legal issues specific to this population (for example, knowing to not refer the enrollees to services that are outside the probationary restrictions). Families First submitted a proposal to a private foundation to fund the COACHES DJJ pilot after the six-month no-cost extension ends in February 2018.

Second, under the new $2 million contract with the state, Families First was expanding its target population to include adolescents ages 15 and 16 while continuing to serve youth ages 17 to 21. Families First expanded the age range to correspond with the eligibility criteria for a foster care independent living program and because, during focus groups, COACHES participants suggested that providing coaching earlier could help keep youth enrolled for at least 12 to 18 months. Families First expected to give COACHES staff training on how to work with younger adolescents.

Replicability. Families First said it was interested in training providers in other states to replicate the program, but it had not yet reached out to them. To make replication easier, Families First trademarked the COACHES name, logo, and implementation protocols.

D. Factors associated with progress toward implementing the SSR plan

Amerigroup's legislative connections and Families First's implementation protocols facilitated SSR of the COACHES program. As the sole managed care organization for foster care youth in Georgia, Amerigroup had developed close relationships with state legislators and staff

at state agencies. Amerigroup noted that its ability to "get in front of the right people to justify our sustainability plan" was a key factor in the awardee's securing the state contract for Families First: "We got in front of the governor's advocacy board [and showed] them our impact [on] graduation rates and grade progression, different health-related measures that we reached, goals that we reached, to really sustain the program." Families First also indicated that developing formal implementation protocols and trademarking the program will help it sustain, scale, and replicate the program while also retaining control over program integrity.

Amerigroup and Families First identified two main barriers to SSR: (1) limited data on program outcomes and (2) slow legislative processes. First, program leaders found it challenging to have enough data and results on outcomes in hand by the time they needed to start asking for sustainability funding. However, they noted that developing fact sheets based on focus group data and early Medicaid outcomes data helped them get buy-in and support from legislators. Second, the legislative process required to finalize the contract between Families First and the state was slow. Families First indicated that "gently pushing" the state, through regular phone calls and emails, helped it to keep the COACHES program on the legislative agenda and finalize the contract before the cooperative agreement ended.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries as part of HCIA R2. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Avera Health

April 18, 2018

Mynti Hossain (Mathematica Policy Research)
Katie Morrison (Mathematica Policy Research)
Evelyn Li (Mathematica Policy Research)
Robert Schmitz (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Residents in skilled nursing facilities (SNFs) routinely experience long waiting periods to see primary care physicians (PCPs) or geriatricians and, as a result, often endure costly and

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

inconvenient transfers to EDs to receive timely medical attention.[3,4] To address this challenge, Avera Health, a nonprofit integrated health system, used HCIA R2 funding to implement the eLongTermCare (eLTC) program. The eLTC program offered a set of geriatric care and tele-health services to staff and residents in SNFs across the Midwest. Services included (1) staff training and empowerment, (2) tele-health transitional care coordination, and (3) tele-health consults for urgent and specialty care. Avera provided eLTC services out of a centrally staffed tele-health hub in Sioux Falls, South Dakota. Overall, Avera implemented the program in 45 SNFs across eight long-term care provider organizations, including its own.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Avera's eLTC program was designed to help residents in SNFs gain access to timely, resident-centered care, particularly in rural areas that were geographically isolated from PCPs and geriatricians, and thus reduce the incidence of unnecessary transfers to EDs and hospitals.
Major innovation	Avera's innovation was in (1) its method of providing instant, around-the-clock access to care for SNF residents through tele-health equipment and (2) its creation and use of a risk-stratification algorithm for SNF residents.
Program components	Quality improvement, telemedicine, and transitional care coordination
Target population	Patients admitted to any of the 45 SNFs participating in the program
Theory of change/theory of action	Avera hypothesized that by providing tele-health services to SNFs and their residents the program would better meet participants' medical needs and in turn reduce ED visits and hospitalizations, both of which would reduce the total cost of care.
Payment model	Capitated payment for services (a retail subscription model—that is, a business model in which SNFs make advance payments to use eLTC services for a specified month)
Award amount	$8,827,572
Effective launch date	November 1, 2014
Program setting	SNFs (provider based)
Market area	Rural, urban, suburban
Market location	IA, MN, NE, and SD
Target outcomes	By August 31, 2017, reduce ED visits by 28 percent, hospitalizations by 16 percent, and the total cost of care by 8.25 percent.

ED = emergency department; PCP = primary care physician; SNF = skilled nursing facility.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the three-year cooperative agreement. We based this conclusion on five factors. First, Avera exceeded

[3] Levy, Cari, Anne Epstein, Lori-Ann Landry, and Andrew Kramer. "Physician Practices in Nursing Homes: Final Report." Washington, DC: U.S. Department of Health and Human Services, April 4, 2006. Available at http://aspe.hhs.gov/daltcp/reports/2006/phypracfr.htm.

[4] Ouslander, Joseph G., and Robert A. Berenson. "Reducing Unnecessary Hospitalizations of Nursing Home Residents." *The New England Journal of Medicine,* vol. 365, 2011, pp. 1165–1167. Available at https://www.nejm.org/doi/full/10.1056/NEJMp1105449#t=article.

both its indirect and direct participant enrollment goals, ultimately enrolling a total of 11,192 indirect participants, or 111 percent of its target indirect participant enrollment goal, by the end of the three-year cooperative agreement. Second, Avera delivered its three services (staff training and empowerment, tele-health transitional care coordination, and tele-health consults for urgent and specialty care) to residents and SNF staff largely as intended throughout all three years of the cooperative agreement. Third, the awardee adequately staffed the eLTC hub 24-7 and met 94 percent of its three-year hiring target. Fourth, although Avera experienced difficulty engaging some PCPs, SNF administrators, and SNF staff, the awardee consistently met with them throughout the cooperative agreement to address their concerns and encourage utilization of eLTC services. Last, although Avera did not focus on engaging program participants because the eLTC program was designed to be an intervention at the SNF level, the awardee collected feedback from residents and their families and addressed any concerns they had.

Impact evaluation. While we do plan to carry out a rigorous assessment of the impacts of Avera's eLTC program, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants. **Payment model.** Avera's payment model is a condition-specific, population-based payment, which SNFs would pay to the awardee in exchange for a bundle of care coordination and clinical services. The awardee has not entered into negotiations with any payers, though it has submitted a proposal for consideration as an alternative payment model to the U.S. Department of Health and Human Services (HHS).

Sustainability plans. Of the 45 SNFs that participated in the cooperative agreement, 33 SNFs agreed to sustain the program through the retail subscription payment model. In addition, all 20 SNFs that participated in the test of the retail subscription model (but not the cooperative agreement) agreed to sustain the program. Avera found the retail subscription model viable for scaling after testing it during the second and third program years. According to the awardee and participating SNFs, the main barrier to sustaining and scaling the eLTC was the program's cost.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of non-clinician staff on their perceptions of the effect of the program on care delivery. The survey—fielded from July 2016 to October 2016 with a sample of 58 SNF staff, including administrators and nurses—achieved a response rate of 72 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of enrollment strategies

Avera employed a three-tier identification, recruitment, and enrollment strategy to secure long-term care provider organizations, their affiliated SNFs, and SNF residents to participate in the eLTC program. First, Avera leveraged existing relationships with long-term care provider organizations that were looking to innovate their services. In states where the awardee was already licensed, Avera invited organizations to enroll in the program. The awardee secured agreements with seven organizations that wanted to participate.[5]

[5] The organizations that agreed to participate were Continuum Health Care Services, Evangelical Lutheran Good Samaritan Society, Skyline Healthcare, Sunrise Retirement, Trinity Health, Veterans Home, and Welcov Healthcare.

Second, to recruit and enroll the organizations' affiliated SNFs, Avera and its partnering organizations targeted SNFs they believed would be most likely to benefit from eLTC program services based on an assessment of the SNFs' hospital admission rates and distance from primary care clinics and hospitals. Avera also conducted an assessment within its own health system to identify Avera-owned SNFs to recruit and enroll in the program. All SNFs that agreed to participate signed contracts that outlined the program requirements.

Third, all residents at the participating SNFs were automatically enrolled in the eLTC program as indirect participants. All residents were considered to be indirect participants because SNF staff who interacted with residents received training through the eLTC program. Indirect participants who received tele-health services paid for by HCIA R2 funds (rather than Medicare, Medicaid, or commercial insurance) were also counted as direct participants.

b. Evidence of enrollment effectiveness

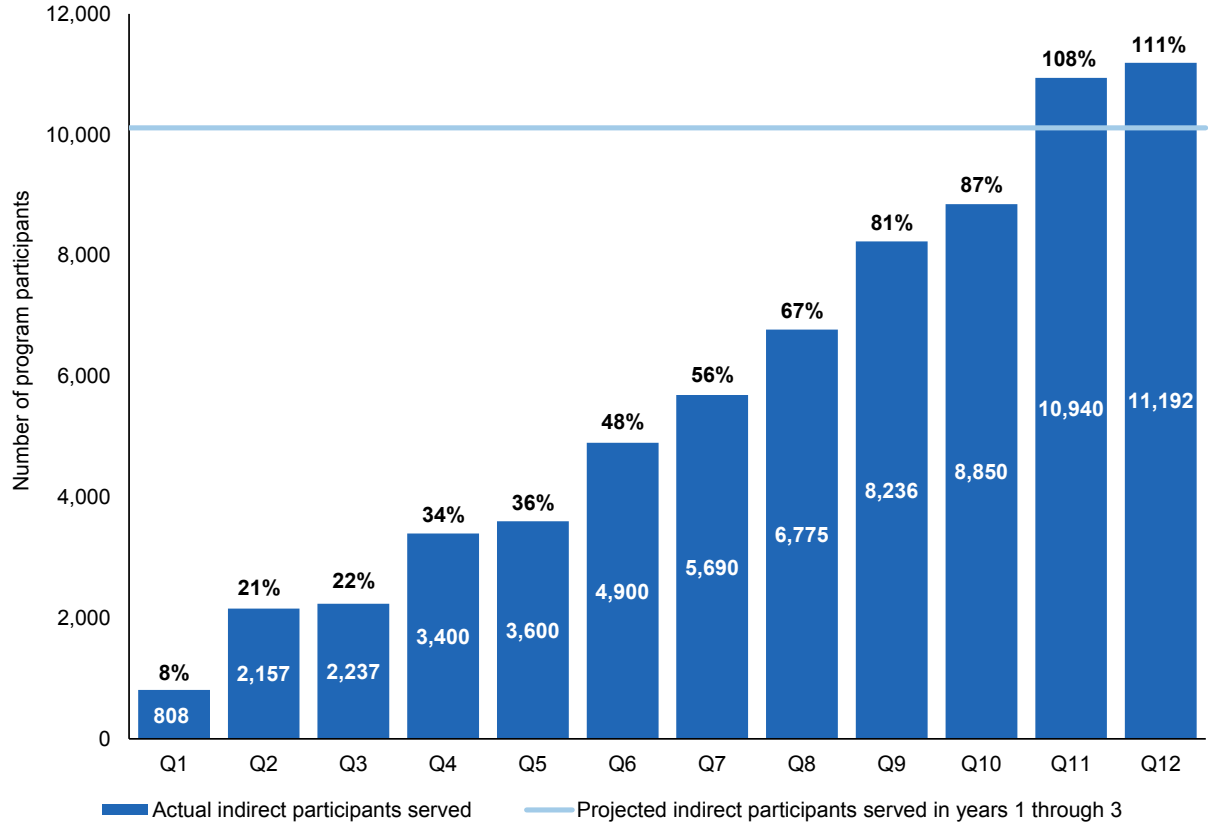
Avera achieved enrollment effectiveness. Overall, Avera reported that it enrolled 11,192 indirect participants from November 2014 through August 2017, which represents 111 percent of its final three-year projection (Figure II.1). Avera originally established a target of 7,100 indirect participants by the end of the three-year cooperative agreement, but increased this projection to 10,115 in the third program year. Avera reported that its projected enrollment goal increased due to an expansion to 15 additional SNFs in the second program year, all of which began using eLTC services in the third program year.[6] Avera enrolled indirect participants at a relatively consistent pace throughout the cooperative agreement. Of the 11,192 indirect participants enrolled by the end of the cooperative agreement, approximately 30 percent were enrolled during the first and second program years and approximately 40 percent were enrolled during the third program year.

Although Avera focused on enrolling indirect participants during the cooperative agreement, it also established a direct participant enrollment goal. Originally, Avera established a direct participant enrollment target of 5,521 across the three-year cooperative agreement. However, Avera decreased this projection to 4,134 in the third program year to more accurately reflect direct participant enrollment in previous months. Nevertheless, Avera exceeded its original goal by serving 6,778 direct participants by the end of the cooperative agreement (Figure II.2).

Skyline Healthcare became a partner during the third program year after the entity purchased 10 SNFs from Golden Living Centers. Before the sale, Golden Living Centers had partnered with Avera to implement the eLTC program in the 10 SNFs.

[6] Each SNF had a different launch date, which Avera defined as the date when the SNFs could call the eLTC hub to request program services. Variation in the date was due to some SNFs requiring less or more time to set up the wireless Internet technology.

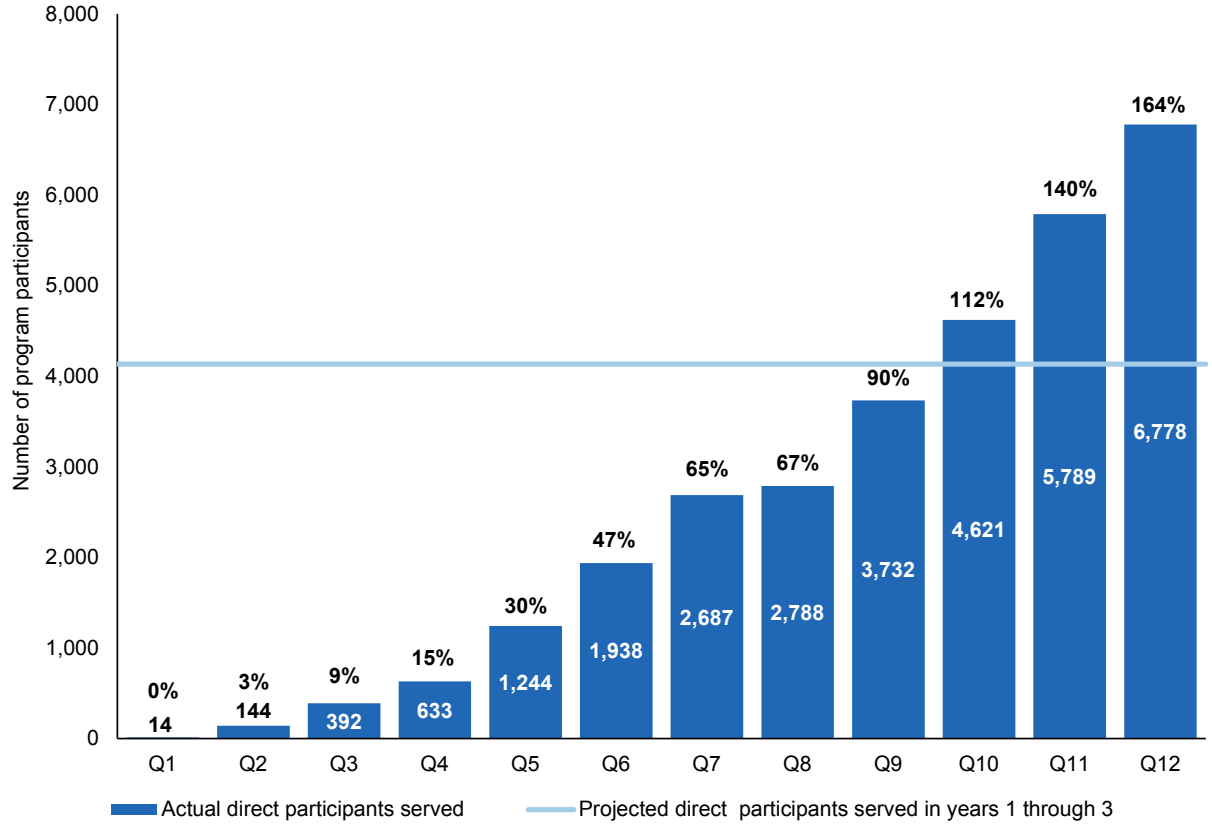
Figure II.1. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Avera considered all enrolled participants (those residing in SNFs) to be indirect participants from program launch through the 12th program quarter because all SNF staff who interacted with residents received training under the cooperative agreement. Although all enrolled program participants were considered to be indirect participants, some indirect participants were also direct participants because they received services not covered under Medicare policy and therefore received tele-health services funded directly by the HCIA R2. All of the direct participants for the eLTC program were also counted as indirect participants. The awardee increased its indirect participant enrollment projection once during the three-year cooperative agreement, from 7,100 at the beginning of the cooperative agreement to 10,115 during the third program year.

Figure II.2. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received tele-health services not covered under Medicare policy (for example, participants residing in urban areas or those in rural areas who received more than one tele-health visit per month). They received eLTC services funded directly by the HCIA R2 from program launch through the 12th program quarter. The awardee decreased its direct participant enrollment projection once during the three-year cooperative agreement, from 5,521 at the beginning of the cooperative agreement to 4,134 during the third program year.

c. Facilitators and barriers associated with enrollment effectiveness

Avera's progress in meeting and exceeding its indirect and direct three-year enrollment goal was influenced by one facilitating factor and three challenges.

The expansion of the eLTC program from 30 SNFs across three organizations to 45 SNFs across eight organizations facilitated enrollment of indirect participants and, in turn, direct participants. Avera was able to carry over funds from the first program year to add 15 additional SNFs to the program in the second year.[7] Given that all residents in participating SNFs were

[7] For more information about Avera's use of carryover funds to expand the eLTC program, see the first annual report at https://downloads.cms.gov/files/cmmi/hcia2-yroneevalrpt.pdf.

automatically enrolled, this meant that there was a greater number of indirect participants served by the eLTC program. This also meant that there was a greater number of indirect participants who were potentially eligible to receive tele-health visits funded by the HCIA R2, and thus be counted as direct participants.

Avera experienced three challenges to direct, but not indirect, enrollment effectiveness: (1) some SNFs experienced difficulty securing the necessary wireless Internet technology to support the intervention; (2) some PCPs, SNF administrators, and SNF staff did not use eLTC tele-health services; and (3) some SNF residents and their caregivers did not want to accept tele-health services. These three factors potentially affected the number of SNF residents who used the eLTC tele-health services, and thus the number of direct participants.

First, some SNFs experienced difficulty securing the wireless Internet technology needed to use eLTC video consults due to wireless dead zones and other technological issues. For example, after Skyline Healthcare purchased 10 SNFs participating in the eLTC program from Golden Living Centers, it experienced difficulty transitioning the SNFs to its network and, in turn, connecting to Avera's network. This led to a three-month period in which those SNFs could not use eLTC video consults. Avera and the Skyline SNFs used telephone consults during the outage and eventually resolved the technological issues. However, the temporary unavailability of eLTC video consults for transitional, urgent, and specialty care slowed the pace of direct participant enrollment at these SNFs during this period.

Second, some PCPs, SNF administrators, and SNF staff were either hesitant or forgot to use the eLTC tele-health services. In addition, some were new to their roles and therefore unfamiliar with the program's services. Throughout the cooperative agreement, some PCPs had reservations about the program because they viewed Avera as a competitor that would potentially interfere with their patient relationships. In the survey of SNF administrators and staff during the beginning of the third program year, more than half of respondents (68 percent) said that clinician resistance to the program was a barrier to meeting program goals. In interviews, some SNF administrators and staff said that they did not use tele-health services because they had work habits that were difficult to change or they were newly hired and thus not aware of the program's services. In addition, Avera reported that job insecurity among staff at the 10 SNFs sold to Skyline exacerbated engagement challenges.

To promote provider engagement with the program and increase their use of its services, Avera conducted in-person and virtual meetings with PCPs, SNF administrators, and SNF nurses on an ongoing basis throughout the cooperative agreement. In the second program year, Avera created virtual training materials for new SNF staff to ensure that all staff were fully trained on when and how to initiate tele-health consults despite high levels of SNF staff turnover. In the third program year, Avera hired an account executive to reach out to PCPs and SNF staff to further encourage them to increase their use of tele-health services. In addition, Avera changed the calculation for the SNFs' third program year incentive payments to reflect their utilization (or lack of utilization) of telephone and video consults, hoping that tying larger incentive payments to higher utilization would spur SNFs' engagement.

Third, Avera reported that some SNF residents and their caregivers were hesitant to accept tele-health services because they were concerned about co-pays. To address this, Avera stopped

billing for tele-health visits in the second program year so that all eLTC program services were free for SNF residents. The awardee also started hosting open houses at the SNFs to promote the program among SNF residents and their caregivers, as well as among SNF staff.

2. Delivery of program services

a. Description of service delivery model

Avera's eLTC program provided three services: (1) staff training and empowerment, (2) tele-health transitional care coordination, and (3) tele-health consults for urgent and specialty care. Avera did not make major changes to the three eLTC services during the cooperative agreement.

Overall, Avera hypothesized that SNF staff would attend the eLTC trainings, integrate information learned into their resident care processes, and reach out to eLTC providers at the tele-health hub for tele-health services as needed. Avera also assumed that eLTC providers at the tele-health hub would have the knowledge, expertise, and tools necessary to (1) provide tele-health transitional care coordination to new SNF residents and (2) respond to urgent and specialty care tele-health requests for SNF residents. The awardee believed that the program would lead to SNF staff using their new knowledge and skills to provide higher quality care, including better identifying situations that were truly emergent and contacting eLTC providers for non-emergent resident care issues. Avera thought these SNF staff behaviors would in turn lead to resident care being delivered more frequently at the SNF rather than at an ED or hospital. Avera viewed the three eLTC services as equally important in reducing ED visits, hospitalizations, and total cost of care among SNF residents. We describe each of the three eLTC services below.

1. **Avera trained and empowered SNF staff.** On a monthly basis, Avera conducted hour-long trainings for all SNF staff, either in person or virtually. Avera selected the topics based on program monitoring data and feedback from SNF staff at each participating site. Avera also provided informal training for SNF staff through ad hoc calls and meetings.[8] In addition, during the first program year, Avera provided support to SNF staff as the SNFs adopted tools from the Interventions to Reduce Acute Care Transfers (INTERACT) quality improvement program.[9]
2. **Avera provided tele-health transitional care coordination.** Avera provided transitional care coordination through a two-step process for newly admitted residents who were transferred from home, a hospital, or other care setting to a participating SNF. First, eLTC staff conducted risk stratification of new SNF residents to differentiate between those who

[8] eLTC providers conducted informal training on advance care planning over the telephone at a SNF staff member's request. In addition, in the second program year, Avera began holding monthly meetings with SNF staff to review any unplanned transfers to an ED or hospital that SNFs made without first consulting eLTC providers. During these calls, Avera trained SNF staff on how to identify (1) early changes in the residents' conditions in order to avoid ED visits or hospitalizations and (2) situations that were truly emergent.

[9] For more information on INTERACT, see https://interact2.net/index.aspx.

were at high risk and low risk for ED visits and hospitalization.[10] Avera initially planned to use the LACE index to stratify residents but found that this algorithm classified almost all residents as high risk.[11] As a result, Avera manually conducted comprehensive reviews of all new residents to determine their risk level, while simultaneously developing its own risk stratification algorithm.[12] Second, depending upon a resident's risk level, eLTC providers delivered additional supports designed to improve care coordination. For high-risk participants, eLTC providers conducted a full geriatric evaluation and developed a tailored ePlan, which included a chronic disease management plan, a schedule for telephone and video consults, and a task list for SNF staff to follow as appropriate. For low-risk participants, eLTC staff reviewed medication lists and provided medication recommendations to the PCP or a SNF nurse. Upon the request of SNF staff, low-risk participants could receive a video consult or an ePlan.

3. **Avera offered tele-health consults for urgent or specialty care.** Avera provided around-the-clock tele-health consults for SNF residents every day. Avera encouraged SNF staff to call providers at the eLTC hub whenever a resident needed urgent medical care. eLTC providers then evaluated the participant via direct two-way audio and video, if necessary, and provided instructions to SNF nurses on next steps for participant care. For non-urgent specialty care, eLTC staff worked with Avera specialists to schedule tele-health visits.

Although Avera had experience providing tele-health services to providers and patients in various care settings, providing these services specifically to SNF staff and resident populations was new to the awardee. Using HCIA R2 funding, Avera expanded eCare, its existing tele-health model, which already included tele-health service lines such as ePharmacy, eEmergency, and eUrgent Care.[13]

b. Evidence of service delivery effectiveness

Avera achieved service delivery effectiveness in four domains. First, Avera delivered its services to residents and SNF staff largely as intended throughout the cooperative agreement. Second, the awardee adequately staffed the eLTC hub around-the-clock and ensured eLTC providers were trained to deliver high quality care. Third, the awardee consistently met with PCPs, SNF administrators, and SNF staff throughout the cooperative agreement to address

[10] In addition to risk stratifying new residents, Avera initially planned to risk stratify all long-term SNF residents on an annual basis. The awardee changed this plan in Year 1 because it found that long-term residents were coming to eLTC providers' attention through the tele-health consults for urgent and specialty care.

[11] LACE is a numeric index that measures the risk of a resident's hospital readmission. The index is calculated from characteristics of the resident and his or her most recent hospitalization. See van Walraven, Carl, Irfan A. Dhalla, Chaim Bell, Edward Etchells, Ian G. Stiell, Kelly Zarnke, Peter C. Austin, and Alan J. Forster. "Derivation and Validation of an Index to Predict Early Death or Unplanned Readmission After Discharge from Hospital to the Community." *Canadian Medical Association Journal,* vol. 182, no. 6, April 6, 2010, pp. 551–557.

[12] In May 2017, Avera deployed a new version of the risk algorithm. The algorithm included 37 elements, most of which were conditions that could increase risk—for example, cancer diagnosis, kidney disease dialysis, and tube feeding. The algorithm categorized residents into categories from 0 (lowest risk for readmission) to 16 (highest risk for readmission). As of July 2017, high risk was defined as a score of 6 or above.

[13] For more information about eCare, see https://www.averaecare.org/ecare/.

concerns and to promote engagement with and use of the eLTC program. Fourth, though the program was not focused on participant engagement, Avera collected feedback from residents and their caregivers and addressed their concerns.

Delivery of intervention services. Avera delivered its three services to residents and SNF staff largely as intended throughout the cooperative agreement. First, Avera provided training to 2,167 SNF staff, which was approximately 105 percent of its three-year training goal.[14] In the survey of SNF staff, 95 percent of respondents reported attending formal eLTC training. Avera reported that it provided more trainings to SNF administrators and staff than anticipated due to a high level of SNF employee turnover as well as a knowledge gap in advance care planning among SNF nurses.

Avera also delivered its tele-health transitional care coordination service largely as planned. In interviews during Year 3, Avera reported that throughout the cooperative agreement most new SNF residents were risk stratified and had received transitional care coordination. In fact, the percentage of new residents across the 45 SNFs who had their medical information (for example, medical charts and medication lists) evaluated by eLTC staff or who received a tele-health evaluation from eLTC providers within 30 days of a SNF admission ranged from 89 percent to 96 percent on a monthly basis from June 2016 to August 2017.

Finally, Avera provided tele-health consults for urgent and specialty care largely as planned. Throughout the cooperative agreement, the awardee repeatedly stated in interviews that it delivered the urgent care consults as intended in regard to length and quality. The length of consults varied depending upon each resident's medical needs. The quality of the consults varied across sites depending upon whether eLTC providers and staff had access to the SNFs' electronic medical records. Avera anticipated this variation in quality and began to address it in Year 2 with the introduction of eLITE, a software program that facilitated the exchange of resident information between the eLTC hub and the SNFs.

Staffing and training. Avera reported that it had hired enough staff at the eLTC hub to implement and manage the eLTC program 24-7 throughout the cooperative agreement. It successfully hired 18 full-time equivalents (FTE), which was 94 percent of its hiring target. In interviews, Avera noted that it had experienced some staff turnover and difficulty with providing timely care during busy periods, such as flu season.

Recruitment and engagement of providers. Avera worked to obtain and maintain buy-in for the eLTC program with PCPs, SNF administrators, and SNF staff throughout the cooperative agreement. Provider engagement was inconsistent across the SNFs. The awardee reported in interviews that SNF staff typically initiated urgent care consults at the appropriate times (that is, when the alternative would have been to transfer a resident to the ED or hospital), with the exception of some SNF staff who were not engaged with the program and did not consistently initiate tele-health consults for their residents. According to its service delivery model, Avera aimed to be contacted by SNF staff before half of all resident transfers to the ED or hospital. The awardee experienced lower-than-anticipated provider engagement and reported that it was only

[14] SNF staff were able to participate in more than one training and, therefore, were counted more than once.

contacted before approximately one quarter of resident transfers during the cooperative agreement.[15]

Engagement of program participants. Avera designed the eLTC program to be a SNF-level intervention directed at SNF staff. It passively enrolled residents at the 45 participating SNFs and did not focus on engaging them as program participants. Nevertheless, the awardee actively sought and responded to participant feedback and reported high levels of participant satisfaction with the eLTC program. In its 12th quarterly report to CMS, Avera reported that 99 percent of residents and their caregivers responding to a survey after a tele-health encounter said they would recommend the eLTC program to others.

In interviews, Avera noted that some residents refused tele-health consults, but said that this happened infrequently. Similarly, in the survey of SNF staff, 69 percent said that resident resistance to the program was not a barrier to program implementation.

c. Facilitators and barriers associated with service delivery effectiveness

Avera's ability to achieve service delivery implementation effectiveness was influenced by four key factors. Three factors facilitated service delivery: (1) previous experience with implementing tele-health programs similar to the eLTC program, (2) value-based payment policies, and (3) the quality of support delivered by eLTC providers. Staff turnover and difficulty with providing timely care during busy periods limited service delivery effectiveness, although the awardee began restructuring its staffing model to overcome these challenges.

First, Avera's experience with implementing multiple tele-health service lines under eCare facilitated implementation of the eLTC program. According to the SNFs and Avera, there were few technological problems once the wireless Internet technology and tele-health equipment was installed. Avera's previous experience with using this technology and equipment allowed it to mitigate any technology and equipment challenges.

Second, value-based payment policies for hospitals facilitated utilization of the program by the SNFs. Hospitals with excess readmissions face penalties and payment reductions under Medicare's Hospital Readmissions and Reduction Program. Hospitals therefore encouraged the SNFs to use eLTC services to reduce readmissions. For some SNFs, this resulted in top-down encouragement from SNF corporate leaders to SNF frontline staff to initiate tele-health consults when residents had urgent care needs—which enabled providers at the eLTC hub to deliver tele-health services as intended.

> "There have been different times where the nurse is going, Something doesn't seem quite right. So they call eLTC and they brainstorm with them: Maybe we should try this or this and then let's see if we need to get physician intervention."
>
> —SNF administrator

Third, the quality of support that eLTC providers delivered to frontline SNF staff facilitated engagement of SNF staff and encouraged utilization of the tele-health consults. Some SNF administrators and

[15] Avera reported on its rate of involvement in resident transfers as part of its self-monitoring and measurement reporting to CMS. To calculate the rate, Avera used the total number of unplanned transfers that it was involved in as the numerator and the total number of unplanned transfers that SNFs made to the ED or hospital as the denominator.

nurses said they valued the support that eLTC providers offered to SNF nurses when they could not work directly with PCPs. In addition, they reported that the eLTC providers treated the nurses as care team partners and encouraged them to also call for non-urgent general help with residents and mentoring. In the survey of SNF staff, 84 percent said that the program made their jobs easier.

C. Assessment of perceived program effects on the delivery of care and outcomes

The majority of staff who responded to the survey at the start of the third program year believed that the eLTC program had positive impacts on the delivery of care and health outcomes (Table II.2). For example, 98 percent of staff reported that the program had a positive impact on the quality of care and services provided to SNF residents and 86 percent said that care coordination was improved because of the program.

> "Last I heard from a resident family, it was $400 for an ambulance…. That's just the ambulance. What about the doctor in the ED, the emergency room, the supplies? To me, saving one person from going to the ED is like [equal to] two or three visits using eLTC."
>
> —SNF administrator

In addition, SNF and Avera staff perceived that the program led to the targeted outcomes of reducing ED visits and hospitalizations and, in turn, the total cost of care. For example, the majority of SNF staff surveyed perceived that the program had reduced ED visits (83 percent) and hospitalizations (88 percent). In addition, in its 12th quarterly report to CMS, Avera reported that 89 percent (the median percentage across all participating SNFs from November 2014 to August 2017) of urgent care video consults resulted in care being provided at a SNF, which averted unnecessary transfers to the ED and hospital. Avera also reported in interviews during the third program year that it believed the program led to reductions in ED visits, hospitalizations, and cost of care based on findings from an internal evaluation of claims data.

Table II.2. SNF staff perceptions of the impact of Avera's eLTC program

The eLTC program had a positive impact on:	Percentage of staff respondents (N = 42)
Quality of care and services provided to participants	98
Ability to respond to participants in a timely way	95
Access to care and services for participants	95
Efficiency of care and services provided to participants	95
Ability to provide care and services that are responsive to participants' preferences, needs, and values	95
Participant quality of life	88
Reduction in the number of hospitalizations	88
Care coordination	86
Reduction in the number of ED visits	83
Reduction in the number of office visits	79

Source: HCIA R2 staff survey, 2016.

ED = emergency department; SNF = skilled nursing facility.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

First, given that all residents in participating SNFs were considered to benefit from the program (via improved care provided by SNF staff after eLTC trainings), it is possible that impacts on participant outcomes may be seen for all participants. Regardless, about 40 percent of Avera's projected indirect participants joined the program in the third program year. These participants may have benefitted from improved care from SNF staff but had less time to do so than participants who joined the program in the first or second program years. Thus, they could dilute overall program impacts. It is also possible that participants who joined late may not have experienced medical needs requiring tele-health supports during the limited time they were enrolled. Therefore, impacts on participants who joined in the third program year may only be seen among those who received tele-health transitional care coordination or urgent or specialty care consults during their period of enrollment in the program. For these reasons, any impact analysis should estimate the impact given participants' time of enrollment and duration of participation.

Second, Avera's difficulty with fully engaging the PCPs, SNF administrators, and SNF staff in the eLTC program may have limited the program's ability to achieve desired outcomes. Avera did not meet its goal to have eLTC providers involved in half of unplanned transfers to the ED and hospital. In some instances, PCPs, SNF administrators, and SNF staff continued to transfer residents to the ED and hospital without first consulting eLTC providers at the tele-health hub, which may have resulted in unnecessary ED visits and hospitalizations.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter provides an update to our assessment of the evaluability of Avera's eLTC program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Avera

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	5,444[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,220
Likelihood of all-cause hospitalizations	763
MDE sample size requirement to detect 20% effect	
Total expenditures	305
Likelihood of all-cause hospitalizations	191
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	Some issues, but probably surmountable; expect to select a comparison group
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We plan to construct two comparison groups from residents of matched nursing facilities in Iowa, South Dakota, and Minnesota—one for newly admitted SNF patients and one for long-term residents of nursing facilities.

B. Characteristics of Medicare and Medicaid participants at baseline

In the previous annual report, we used facility admission and discharge data from Avera to identify the treatment group. For this annual report, we used administrative data to identify the treatment group beneficiaries because to estimate the impact of the intervention we must use the same data source to identify the comparison group beneficiaries. When we used Minimum Data Set 3.0 (MDS) to identify the treatment group beneficiaries, we found about 27 percent more treatment beneficiaries than we did when using the awardee's enrollment finder file.[16] This discrepancy led us to use data from the MDS to identify treatment group beneficiaries. Therefore, we will use the same data source to construct both the treatment and comparison group samples to eliminate any potential bias due to the differences in the sample identification process.

As with previous reports and consistent with Avera's enrollment criteria, all Medicare beneficiaries who were residing in a participating facility at the start of Avera's eLTC program or who became residents of participating facilities after the start of the program were eligible for the treatment population. Facilities began participating in the program between November 2014 and October 2015. To identify the treatment group by using the MDS, we required that the individual resided in an Avera facility during the intervention period. We were able to attribute treatment beneficiaries based on their facility because the intervention was implemented at the facility level; thus, all residents at a participating facility could potentially benefit from the program.

Consistent with previous reports, a participant's enrollment date depends upon his or her status as an existing or new resident in a participating facility on or after the program start date. The program enrollment date for an existing resident at the program start-up is the date on which the program began at his or her facility; the enrollment date for a new resident is the first day on which he or she became a resident in a participating facility. Yankton Care Center in South Dakota, for example, began the eLTC program on November 1, 2014. We assigned November 1, 2014, as the enrollment date for residents of the facility on that date and assigned the admission date as the enrollment date for residents who were admitted to the facility after November 1, 2014.

We applied specific eligibility criteria for Medicare beneficiaries in order to present baseline characteristics and health care utilization outcomes in this report. First, we included in the treatment group beneficiaries who were residents of participating facilities during the program period if they were enrolled in Medicare fee-for-service (FFS), both Parts A and B, with Medicare as the primary payer on their program enrollment date. Second, beneficiaries would have to meet the first criterion for a period of at least 90 days during the baseline year (the 365 days immediately before their enrollment date). Third, beneficiaries had to be residents of a

[16] A large proportion (about 20 percent) of the beneficiaries in the Avera finder files lacked valid Social Security numbers, dates of birth, or HIC numbers, so we could not verify their Medicare FFS status.

participating facility on or before November 30, 2016, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants.

As of November 30, 2016, we identified 4,852 Medicare beneficiaries in the MDS data who met the eligibility criteria for the Avera eLTC program[17]. We refer to these beneficiaries as the treatment group for the evaluation. The baseline demographic and health status characteristics in this report include information for these individuals (Table III.2). Given the program's focus on long-term care, Avera program participants have substantially poorer health and greater care needs than most Medicare FFS beneficiaries. Most participants are age 85 or older (46 percent) or 75 to 85 years old (30 percent). Most participants are female (65 percent) and white (93 percent). For the majority of participants (80 percent), the original reason for Medicare eligibility was age (65 or older). For the remaining participants, the original reason for Medicare eligibility was disability (19 percent) or end-stage renal disease (ESRD) (fewer than 1 percent). Twenty-three percent of participants are dually eligible for Medicare and Medicaid, although in some cases Medicaid benefits may be restricted to the payment of co-insurance and deductibles. The average HCC risk score of participants (2.46) is nearly two and a half times higher than the average for Medicare FFS beneficiaries nationally (1.0). More than three-quarters of the participants had HCC risk scores higher than the national average.

Participants also had high rates of service use and Medicare expenditures in the year before enrollment. In Table III.3, we report baseline utilization and expenditure data for a common set of measures. We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $2,245—roughly three times the U.S. average Medicare expenditure per enrollee in 2014. Average PBPM Medicare payments for inpatient ($1,011); SNF ($345); and outpatient ($387) services were the largest drivers of total cost of care.

The rates of acute care hospitalizations and ED visits were moderate for this institutionalized population of Medicare beneficiaries. The rate of acute care hospitalizations was 1,062 per 1,000 Medicare FFS participants per year during the baseline year—more than four times higher than the U.S. average of 274 per 1,000 individuals in 2014.[18] The rate of ED visits not leading to hospitalization was 952 per 1,000 participants per year in the baseline year. These high rates of service use suggest opportunities to reduce avoidable hospitalizations and ED visits through

[17] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

[18] The PBPM expenditure as well as the hospitalization and ED visit rates were drawn from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html.

enhanced care coordination. The rate of primary care visits was 9,603 per 1,000 Medicare FFS participants per year, while the rate of specialty services was 10,620 per 1,000 participants per year. The likelihood of a 30-day readmission was close to the national average, at 18 percent per discharge and 7 percent per beneficiary. Utilization and expenditures for nearly all services were considerably higher in the baseline fourth quarter compared with the first through third quarters. This pattern is consistent with the expectation that many new residents likely had costly acute care services immediately prior to the nursing home admission that triggered their program enrollment.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Avera's program through November 30, 2016

	All participants (N = 4,852)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	387	8
65 to 74	751	15
75 to 84	1,460	30
85 and older	2,254	46
Gender		
Female	3,159	65
Male	1,693	35
Race		
White	4,527	93
Black	8	0.16
American Indian, Alaska Native, Asian/Pacific Island American, or other	307	6
Hispanic	4	0.08
Original reason for Medicare eligibility		
Old age and survivor's insurance	3,897	80
Disability insurance benefits	907	19
ESRD[a]	48	0.99
Hospice[b]	89	2
Medicare/Medicaid dual status, percentage dual[b]	1,107	23
HCC score[c]		Statistic
Mean		2.46
25th percentile		1.31
Median		2
75th percentile		3.16

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date for existing residents is the date that the program began at their facility. The enrollment date for new

Table III.2 *(continued)*

residents is their first day of residence in a participating facility. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

ESRD = end-stage renal disease; FFS= fee-for-service; HCC = hierarchical condition category.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Avera's program through November 30, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	4,852	4,801	4,825	4,849	4,852
Average Medicare expenditures PBPM[a]					
Total	2,245 (44)	1,216 (62)	1,311 (44)	1,519 (64)	4,907 (103)
Acute inpatient	1,011 (26)	347 (20)	376 (22)	499 (39)	2,803 (69)
Inpatient other[b]	99 (10)	33 (7)	37 (8)	73 (19)	250 (27)
Outpatient[c]	387 (10)	322 (12)	350 (13)	355 (12)	520 (14)
Physician services	296 (7)	189 (7)	201 (7)	211 (7)	580 (12)
Home health	39 (2)	37 (3)	31 (3)	42 (4)	46 (3)
Skilled nursing facility	345 (13)	234 (50)	254 (18)	266 (21)	625 (35)
Hospice	35 (5)	21 (4)	29 (5)	40 (7)	50 (6)
Durable medical equipment	33 (1)	32 (2)	33 (2)	32 (2)	33 (2)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	1,062 (18)	478 (23)	491 (23)	565 (31)	2,696 (43)
Outpatient ED visits[e]	952 (23)	744 (31)	737 (33)	748 (32)	1,573 (45)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Primary care visits in any setting	9,603 (129)	6,899 (153)	7,212 (155)	7,562 (158)	16,674 (269)
Primary care visits in ambulatory settings	5,713 (77)	5,159 (96)	5,366 (96)	5,482 (96)	6,834 (112)
Specialist visits in any setting	10,620 (207)	7,837 (217)	7,814 (204)	8,321 (225)	18,430 (433)
Specialist visits in ambulatory settings	6,080 (120)	5,829 (150)	5,782 (137)	5,931 (142)	6,767 (141)
Measures of any health care utilization					
Percentage with a hospital admission[d]	65 (1)	10 (<0.5)	10 (<0.5)	11 (<0.5)	54 (1)
Percentage with an outpatient ED visit[e]	48 (1)	14 (1)	13 (<0.5)	14 (1)	27 (1)
Percentage with a 30-day readmission among all discharges	18 (1)	13 (1)	12 (1)	16 (1)	25 (1)
Percentage of participants with a readmission among all participants	7 (<0.5)	1 (<0.5)	1 (<0.5)	2 (<0.5)	3 (<0.5)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Avera's payment model is a condition-specific, population-based payment that SNFs would pay to the awardee in exchange for a bundle of care coordination and clinical services. The awardee has not entered into any negotiations with payers, though it has submitted a proposal for consideration as an alternative payment model to HHS.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

The awardee describes its condition-specific, population-based payment as a retail subscription model in which SNFs would pay Avera to use eLTC services. The incentive for SNFs to participate in this program is to demonstrate fewer ED visits and hospital admissions, which can position SNFs to partner with entities such as accountable care organizations (ACOs) that are pursuing value-based purchasing arrangements with financial incentives for lowering Medicare costs.

For every admitted resident, SNFs pay Avera a fee at the time of admission and then a monthly fee until the resident is discharged. The fee is paid for all residents included in a SNF's monthly census, regardless of the resident's payer or length of stay. Services covered include SNF staff training and empowerment, tele-health transitional care coordination, and urgent and specialty care consults via Avera's tele-health hub. Avera anticipates that these three eLTC services will help SNFs maintain or improve their performance on selected measures used in CMS's Nursing Home Compare and Skilled Nursing Facility Value-Based Purchasing program. If the SNFs do not maintain or improve their performance, then the fee they pay to Avera could be reduced by up to 50 percent beginning in the second year of program participation.

To assess the effect of the eLTC program on Medicare costs, Avera conducted analyses by using Medicare claims data for beneficiaries in participating SNFs. Avera estimated that total Part A and B costs decreased $342 PBPM among the SNFs compared to their historical Part A and B spending—which was defined as the three years prior to participation in the program. The awardee has not been able to compare these costs to non-participating SNFs nor is it aware of any return on investment analyses conducted by individual SNFs.

Avera has developed an alternative payment model, referred to as a performance-based model, that is a variant of the retail subscription model. The main difference is that rather than having a SNF pay a monthly fee to Avera for participating in the program, a member of Avera's eLTC provider team would use his or her national provider number to bill Medicare for services rendered to Medicare beneficiaries served by the SNF. Avera estimates that this represents approximately 90 percent of a SNF's case mix.

A second payment model developed by Avera would incorporate shared savings into the performance-based model. Similar to CMS's Bundled Payment for Care Improvements, there would be a "target bundle experience," based off historical Part A and B costs for similar admissions from the time of admission to the SNF to 30 days post-discharge. The target bundle experience would be compared to the actual spending for all Part A and B costs from the time of SNF admission to 30 days post-discharge. The difference between the target bundle experience and the actual experience would be the shared savings. If the actual experience was less than the target bundle experience, then 50 percent of savings would be shared with the billing entity (Avera, in this case). If the actual experience was greater than the target bundle experience, then the billing entity (after an initial period of upside risk only) would be responsible for repaying Medicare up to 50 percent of these costs. Repayment would be made through a formula that accounts for certain performance criteria and for the number of years in the shared savings program. Avera anticipates that the shared savings option would apply to Medicare beneficiaries who are not already attributed to an ACO.

C. Status of the payment model

At the time of our interview, Avera was not in discussions with payers about its retail subscription model and its program services were not reimbursed by any payers. However, 53 of the 65 SNFs that participated in either the cooperative agreement or the retail subscription test planned to continue paying Avera for program services—which the SNFs viewed as valuable in reducing hospital admissions.

D. Factors associated with the development of the payment model

In previous rounds of data collection, the awardee noted the challenges of developing a payment model within a value-based payment environment, especially among ACOs and Medicare Advantage plans. These organizations were often already focused on shared savings in long-term care.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Of the 45 SNFs that participated in the eLTC program under the cooperative agreement, 33 SNFs agreed to sustain the program through the retail subscription payment model. In addition, all 20 SNFs that participated in the test of the retail subscription model (but not the cooperative agreement) agreed to sustain the program. Avera found the retail subscription model viable for scaling after testing it during the second and third program years. According to the awardee and participating SNFs, the main barrier to sustaining and scaling the program was the cost.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In the second program year, Avera started developing and testing payment models to sustain and scale the eLTC program at current and new SNFs. The awardee had begun developing the retail subscription model, which would charge implementing sites fees for eLTC services after the cooperative agreement ended. The awardee also determined that SNFs could potentially support the program by billing current procedural terminology codes for chronic care management and transitional care management—pending modifications to current codes as well as regulatory removal of tele-health rural designation requirements and limitations to the number

of tele-health encounters that can be provided. Avera began testing the retail subscription model at 20 SNFs that were not part of the cooperative agreement. By the second program year, the awardee reported that 16 SNFs had finalized agreements with Avera to sustain the program. Once tested, Avera reported plans to market the program to more SNFs nationwide.

C. Implementing the SSR plan: progress and changes

Sustainability. As of the third program year, Avera reported that 53 of the 65 SNFs that had participated in either the cooperative agreement or the payment model test would sustain the program through the retail subscription model. Avera leaders met with SNFs through the end of the cooperative agreement to discuss sustainability and help develop sustainability plans, if needed. Awardee leaders hoped to sign contracts with all sustaining SNFs by the end of August 2017.

Scalability. In the third program year, Avera leaders reported finding the retail subscription model viable for scaling. Awardee leaders reported that the testing phase determined the existence of market demand for the program and the ability to successfully implement the program in a variety of SNFs.

Replicability. Avera did not report plans to replicate the program.

D. Factors associated with progress toward implementing the SSR plan

According to the awardee, the main challenge for SNFs to sustain and scale the program was the cost. First, the awardee struggled to convince managed care plans and ACOs that the eLTC program yielded enough savings to justify the program's cost. Although health care providers may experience increased value-based payments by improving re-hospitalization scores, the awardee said that other outcomes such as improved quality of care and appropriate care transitions do not easily translate to cost savings. Awardee leaders said that the perceived low cost savings cooled payers' interest in the eLTC program at some sites. One interviewee suggested that SNFs operating in small, rural markets had more difficulty justifying the cost of the program than those in large, urban areas.

> "Savings from improving re-hospitalization scores are difficult to convey as a return on investment indicator. [Similarly], residents experience the benefit of improved quality of care and appropriate care transitions; however, those outcomes are challenging to equate to a financial cost savings."
>
> —Awardee-submitted report

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. However, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or the SSR plans. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

In the coming year, Mathematica will use propensity score matching to select a comparison group that is well-matched to all beneficiaries who enrolled over the entire period in which Avera's program operated. Using this sample of treatment and comparison beneficiaries, we will estimate the models to arrive at final impact estimates for the life of the program and for different intervals of time since patients' enrollment.

We will synthesize these findings across time periods and outcomes to draw inferences about overall program effectiveness in the final report in August 2019.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Boston Medical Center

April 18, 2018

Dana Petersen (Mathematica Policy Research)
Danielle Chelminsky (Mathematica Policy Research)
Anna Christensen (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Boston Medical Center received a four-month extension through December 31, 2017. The awardee ended enrollment in August 2017 and used the remaining months of the extension to complete data analysis and continue sustainability efforts.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

Boston Medical Center and its implementing partner, Baystate Medical Center, are acute care, nonprofit, academic medical centers located in urban communities in Massachusetts. Together, the two sites used HCIA R2 funding to implement the Collaborative Consultative Care Coordination (4C) program to address the needs of children with medical complexity (CMC) (Table I.1). The 4C program, which was launched in December 2014, used multidisciplinary teams to provide comprehensive and personalized care coordination to CMC and their families under the direction of pediatricians who specialize in care to CMC. The program was innovative because it offered a greater intensity of care coordination supports and access to care teams with a broader array of disciplines than what was previously offered at the Boston and Baystate sites. It was also innovative because it offered consultation services to primary care physicians (PCPs), who are often not trained specifically or who do not have the practice resources to support CMC. The use of a care plan, which was co-developed by the care team and the family as the guide for tailored service delivery, was also a key feature of the program. Staff used a secure, cloud-based, Internet portal (ACT.md) to make the care plan available to families, PCPs, and other providers involved in the child's care.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The 4C program provided consultation support to the PCPs of CMC and helped the CMC and their families coordinate social, behavioral, and medical services.
Major innovation	The 4C program was innovative because it offered a greater intensity of care coordination supports and access to care teams with a broader array of disciplines than what was previously available and emphasized consultation with the CMC's PCPs. The use of an Internet-based, shared care plan co-developed by the care team and the family was also a key feature of the program.
Program components	• Care coordination • Health information technology
Target population	The 4C program defined CMC as children who were diagnosed with at least one chronic medical condition in any of nine categories[a] and who had high service use in the year before enrollment or who were considered by the 4C staff to be at risk for high service use. CMC could enroll in the 4C program regardless of insurance status or type.
Theory of change/ theory of action	Boston Medical Center hypothesized that improving care coordination for CMC would lead to increased access to appropriate services and supports for CMC and their families, which would decrease caregiver stress and depression and improve the CMC's health. Better CMC health would lead to fewer hospital stays and thus decreased health care costs.
Payment model	Both sites are in discussions with hospitals, payers, and other stakeholders. Each site expects to reach an agreement with a local Medicaid accountable care organization for inclusion beginning in March 2018.
Award amount	$6,128,059
Launch date[b]	December 12, 2014
Program setting	Two acute care, nonprofit, academic medical centers
Market area	Urban

Table I.1 *(continued)*

Program characteristic	Description
Market location	Boston and Springfield, MA
Target outcomes	• Improved care planning and coordination • Less caregiver stress and depression • Lower cost of care related to fewer hospitalizations

[a]The categories are (1) neuromuscular, (2) respiratory, (3) cardiovascular, (4) renal, (5) hematologic, (6) immunologic, (7) metabolic, (8) autism spectrum, and (9) congenital defect.

[b]After the initial planning period, the awardee's program became operational as of this date.

CMC = children with medical complexity; PCP = primary care physician.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we conclude that Boston Medical Center was partly successful at implementing its program by the end of the three-year cooperative agreement. We based this conclusion on six factors. First, Boston Medical Center enrolled 365 participants—81 percent of its enrollment target—by the end of the three-year cooperative agreement. Second, the awardee delivered services as intended and on schedule for most participants, though staff were challenged in later years with providing the same intensity as in the first program year because their caseloads grew. Third, Boston Medical Center recruited, hired, and retained staff, but struggled with providing training and supports that met the staff's needs. Fourth, the awardee had difficulty engaging PCPs and other providers serving CMC in care coordination activities and in using the ACT.md care coordination software. This was especially problematic given that collaboration with PCPs was a key tenant of the 4C service delivery model. Fifth, although the awardee was successful in its initial engagement of participants' families, maintaining their long-term engagement in the program was more challenging, as was encouraging their use of ACT.md. Despite these challenges, program leaders and staff felt that the program had a positive effect on the delivery of care.

Impact evaluation. Due to too few treatment beneficiaries, we do not anticipate being able to conduct a rigorous impact analysis for Boston Medical Center.

Payment model. Both the Boston and Baystate sites were actively participating in discussions with hospitals, payers, and other stakeholders as part of a recent statewide effort to implement regional Medicaid accountable care organizations (ACOs). They expected to reach agreements in preparation for launching the ACOs in March 2018. One of the new ACOs, the Boston Accountable Care Organization with Boston Medical Center Health Net Plan, will include Boston Medical Center, its physician practices, most of its licensed and affiliated community health centers, and other providers. Baystate and its community health centers will join with a managed care organization to form the Baystate Health Care Alliance with Health New England ACO.

Sustainability. Although the awardee was still awaiting the launch of its payment model (Medicaid ACOs), program leaders at both sites (Boston and Baystate) worked to focus the 4C program on its most effective components to help sustain it after the cooperative agreement. Both sites were trying to identify funding and strategies for scaling the program to other

providers who will be involved in the accountable care organization (ACO), as well as for participants who will not be included in the ACO. There appeared to be little activity to replicate the program, although the awardee planned to use part of its no-cost extension to disseminate information about the program throughout the country.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey, which was fielded from July to October 2016 with a sample of 21 4C team members and program administrative staff, achieved a response rate of 95 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items with fewer than 11 responses, we described the findings in qualitative terms to avoid identifying respondents. We did not weight the survey samples to adjust for nonrespondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that Boston Medical Center was partly successful at implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

The Boston Medical Center aimed to actively enroll 450 CMC by the end of the cooperative agreement. The 4C program defined eligible CMC as children who were diagnosed with at least one chronic medical condition in any of nine categories and who had high service use in the year before enrollment or who were considered by the 4C staff to be at risk for high service use. The eligible participant's caregiver and PCP both consented to the child's participation in the program. The awardee invested in multiple outreach and recruitment efforts, including communicating directly with PCPs to identify patients who qualified; meeting with families of potential participants during their hospital, neonatal intensive care unit, or pediatric intensive care unit stays; advertising at community health centers, schools, and social service agencies; and conducting other outreach such as presentations.

The 4C staff followed up on all referrals to assess initial eligibility by using available electronic medical record (EMR) data. The 4C complex care pediatrician or nurse practitioner then connected with the CMC's PCP to gain approval for the child's participation in the program. If the PCP approved and the child met initial eligibility criteria, the 4C program coordinator informed the family about the program, extended an invitation to participate, and scheduled an intake assessment.

The 4C program enrolled participants after they completed the initial intake assessment, were confirmed as eligible, and signed a consent form.[4] Although the eligibility criteria were more rigid at the beginning of the cooperative agreement, over time the 4C team relaxed the criteria and relied more heavily on the clinical judgment of the complex care pediatrician. In the second and third program years, the 4C team generally confirmed eligibility for the program after reviewing each case as a team and applying clinical judgment to make a final enrollment determination. This meant that the program sometimes, though rarely, enrolled participants who did not strictly meet the medical criteria but lived in families experiencing complex social needs. Children could enroll in the 4C program regardless of insurance status or type. Sixty-nine percent of enrolled participants were covered by Medicaid or CHIP.

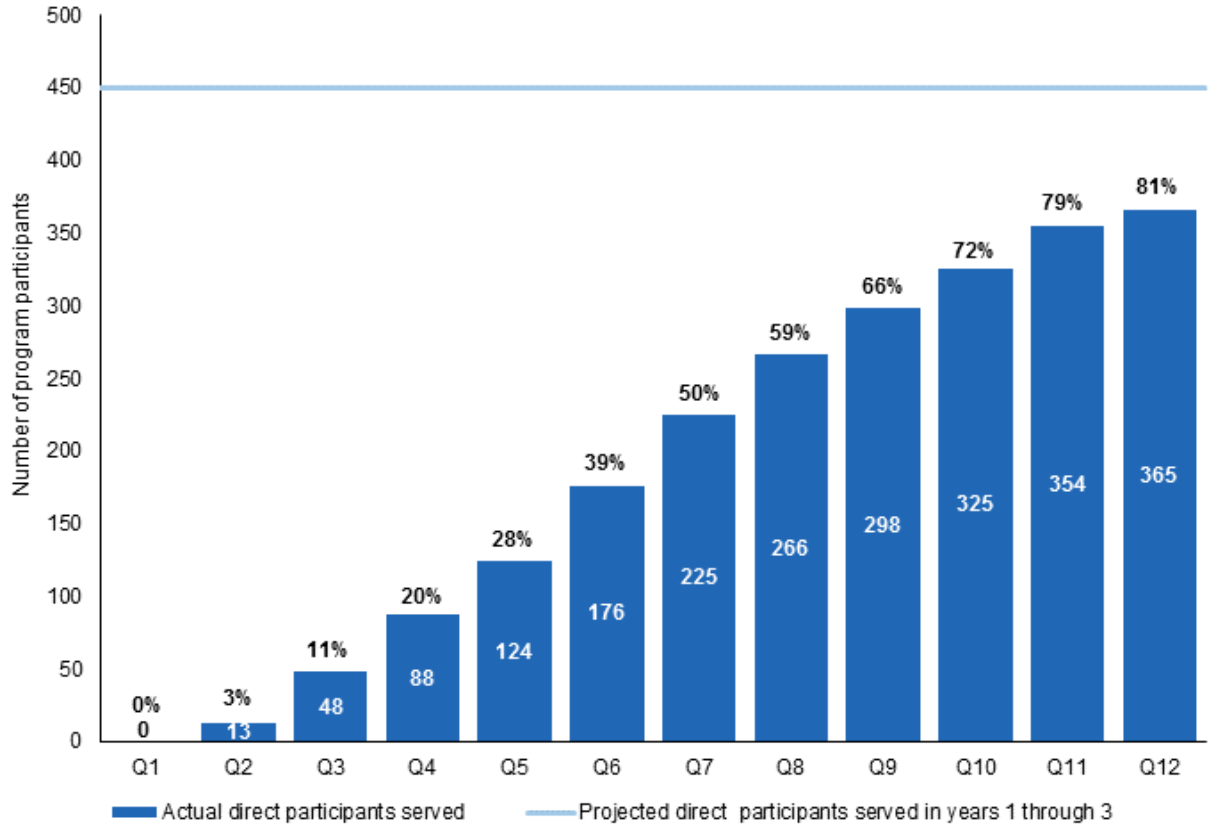
b. Evidence of enrollment effectiveness

After a slow start to enrollment during the first program year and steady progress in the second program year, Boston Medical Center came close to meeting its enrollment goal in the final year of the cooperative agreement. Overall, the awardee was partially effective in achieving its enrollment goal, enrolling 365 participants from December 2014 (when it launched its program) through August 2017. The awardee met 81 percent of its enrollment goal of 450 direct participants despite being unable to realize its initial plan to identify potential participants by using claims information submitted by payers. The awardee was unable to overcome payers' privacy concerns regarding the sharing of necessary claims data (Figure II.1). Boston Medical Center did not change its enrollment targets over the course of the cooperative agreement.

According to program staff, the awardee enrolled participants who matched the intended target population. The Boston site, however, enrolled a higher-than-expected proportion of children with complex behavioral health challenges. Staff also reported that the awardee enrolled most participants early enough in the program to receive a sufficient amount and duration of services to expect improvements in the 4C program's desired outcomes. A few participants were lost to follow-up due to death or unsuccessful contact attempts.

[4] After enrolling, participants remained enrolled for the duration of the program, unless they reached the age of 22. At that point, they were transitioned into adult care.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017)

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee did not change its enrollment target over the course of the cooperative agreement.

c. Barriers and facilitators associated with enrollment effectiveness

Boston Medical Center's progress in meeting its three-year enrollment goal was influenced by several factors. Four factors facilitated the awardee's enrollment effectiveness: (1) the 4C program's reputation, (2) "warm handoffs" to the program by the child's PCP, (3) face-to-face contact between 4C staff and CMC families, and (4) the use EMR data. Two factors hindered the awardee's enrollment efforts: (1) PCP reluctance to refer patients and (2) the challenges of being a parent of a CMC.

> "Word of mouth from patients' families … was very helpful, and the value that PCPs and family members felt they were receiving from the program led to additional referrals."
>
> —Program leader

Staff from both sites reported that their reputation of providing services for CMC that was unmatched by other local providers facilitated enrollment in later program years. That growing reputation resulted in more referrals from providers and families via word of mouth and in more internal referrals from within the respective hospital systems in later program years compared to the first program year. As a result, the awardee did not need to conduct outreach and enrollment

activities as aggressively over time. In addition, staff from both sites noted that having the child's PCP conduct a warm handoff and endorse the program to potential enrollees supported enrollment. The 4C staff continued to report that meeting face-to-face with families, either during home visits or by showing up at their existing appointments with PCPs, was a strong facilitator of enrollment. During the third program year, Boston site staff started using EMR data to identify and recruit participants within their own hospital system, which also supported enrollment efforts.

The factors that hindered 4C program enrollment in prior years continued to do so in Year 3. For example, although the program received more referrals from providers in later years than in the first program year, some PCPs continued to be reluctant to refer patients because they did not fully understand the 4C program model and were fearful that the program was taking their patients away. Baystate and Boston Medical Center leaders and staff also reported that the daunting challenges of being a parent of a CMC posed an obstacle to enrollment. In addition, as in previous years, staff noted in interviews that some parents of CMC were resistant to enrolling their children because they felt overwhelmed with the number of providers and appointments they already had or they were hesitant to change their support systems. In the survey of program staff, about half of staff reported that participants might not enroll or participate because they viewed the program as having too many requirements (such as meetings with the care team) or too much of a time commitment. Although the 4C staff continued to educate and inform parents of the benefits of care coordination, some parents were not convinced.

2. Delivery of program services

a. Description of and changes to service delivery model

Through the 4C program, Boston Medical Center sought to change the way that care to CMC was delivered by providing their families and PCPs with consultation services and care coordination supports guided by a comprehensive care plan. The awardee aimed to address families' medical, social, and behavioral health challenges and barriers to accessing services while also bolstering the confidence and competence of PCPs who treat CMC. The 4C program staff also expected to use ACT.md—a secure, cloud-based, Internet portal—to make comprehensive care plans available to the families, PCPs, and other providers involved in CMC care and to track their care coordination activities. Key components of the 4C service delivery model included the following:

- Building multidisciplinary teams to engage with CMC and their families. Each team was composed of a nurse care coordinator, a social worker, a family navigator, a dietician, a child psychiatrist (or child psychologist), and in some cases a developmental behavioral physician. A complex care pediatrician, who typically served in a consultant role and not as the participant's PCP, oversaw the work of the two teams at each site.
- Conducting multidisciplinary intake assessments of the medical history and health status of CMC, their health care and health care service needs, and the social needs of their families.
- Developing comprehensive care plans with CMC families that outlined the medical and social goals and needs of the children and their families.

- Meeting with CMC and their families for a follow-up one month after the intake assessment and then every six months thereafter. Families and 4C staff also scheduled interval appointments and met or talked on the telephone informally as needed.
- Offering consultation services to PCPs who provide care for CMC.
- Improving communication across the multiple providers serving the CMC and facilitating access to medical and social support services needed by the CMC and their families to achieve their goals.

The 4C program built on prior work conducted through the Pediatric Comprehensive Care Program at Boston Medical Center and the CMC Medical Home and Care Coordination Programs at Baystate Medical Center by offering a greater intensity of care coordination supports and access to care teams that included a broader array of disciplines. For example, the 4C program added family navigators. In addition, the 4C program prioritized outreach and consultation services for PCPs, who often struggle to provide comprehensive, coordinated care for CMC on their patient panels. The vision of the 4C program was that PCPs and families would be matched with a 4C team to coordinate a CMC's medical and allied health services. Finally, the program linked the Boston and Baystate sites, which had not worked together before. This linking allowed for shared learning and problem-solving across sites.

b. Evidence of service delivery effectiveness

Based on our analysis, Boston Medical Center was partially effective in delivering program services. The awardee's successes included initiating service delivery on schedule; initial engagement of participants; and recruiting, hiring, and retaining staff. For the most part, staff reported delivering services as intended. However, the awardee struggled with providing training and supports that met staff needs, engaging PCPs and other providers serving CMC in care coordination activities, engaging participants' families over the long term, and getting both families and providers to use ACT.md to communicate with staff or update the care plans.

Delivery of intervention services. According to 4C leaders and staff, the program was mostly successful in delivering services to CMC and their families as intended. Staff reported performing intake assessments and developing care plans for all enrolled CMC. They reported conducting at least the minimum required follow-ups at one month after intake and every six months thereafter for most participants, though some six-month appointments occurred by phone rather than in person during the third program year. The awardee reported that, as of May 2017, it had provided more than 15,000 unique encounters to its 365 participants since the program's launch. In the last program quarter (between March and May 2017), these encounters included 852 telephone contacts, 266 in-person appointments, and 107 online or email contacts with 4C participants.

During interviews, the nurse care coordinators, family navigators, and social workers described providing care coordination supports consistent with the service delivery model: (1) making referrals to and assisting families with making appointments with other providers, (2) attending appointments together with families, (3) acquiring medical supplies for CMC and teaching families how to use those supplies, (4) assisting with access to special services at schools, (5) finding housing or food supports, (6) helping families overcome transportation

challenges, and (7) helping families to be more self-sufficient in addressing these needs. They also described monitoring families' progress and assigning team members with follow-up tasks when needed. However, staff reported difficulty providing the same level of intensive care coordination in between appointments in the third program year as in the first program year because their caseloads grew. Sixty percent of staff completing the non-clinician survey around the start of the third program year felt that their participant caseload was either much too heavy or somewhat heavy, while 40 percent reported that they had insufficient time for the amount of work they wanted to do.

Staffing and training. Both program sites were successful in recruiting, hiring, and retaining the staff they needed to deliver services. The awardee hired staff and established two care teams at each site, with minor turnover or vacancies. It did not significantly change its staffing model over time. As of May 2017, the awardee reported that it had exceed its hiring goals and had achieved a 94 percent retention rate for staff.

In interviews during the third program year, staff perceptions of the adequacy and effectiveness of training and supports provided by the awardee varied. Although some staff said they liked relying on their own personal experience and judgment when performing the functions of their job, others desired more training, supervision, and guidance. Thirty-five percent of staff responding to the non-clinician survey around the start of the third program year reported that they had attended formal training and 84 percent reported that they had received informal instruction related to the program. Informal training was provided through staff meetings (89 percent); huddles and self-study (80 percent each); asking colleagues for help (75 percent); mentoring, shadowing, or technical assistance (55 percent each); and individual and group supervision (40 percent each).

> "We did have support from local care coordinators within the medical home [program in the community, external to the 4C program]. Initial information was learned that way. But specific to this population [of CMC] and learning to do what we are doing as part of the 4C program, that was kind of a work in progress. You know, learning as we go."
>
> —Nurse care coordinator

> "I thought that [providers] would be enticed by the information once it was housed in ACT.md…. I thought that would get buy-in, but it really hasn't."
>
> —Nurse care coordinator

Recruitment and engagement of providers. According to staff interviews, the program struggled throughout the three-year cooperative agreement to achieve the level of desired collaboration and engagement among PCPs, specialists, and other providers who served the CMC. Staff also struggled to get PCPs and other providers to use ACT.md as a method for communicating or electronically sharing and updating the CMC comprehensive care plans.

Engagement of program participants. The awardee was mostly successful in engaging participants' families. In interviews, staff described some families as being so engaged that they sometimes overwhelmed staff with their continual requests for assistance, while a smaller subset of families was hard to engage over the long term. Eighty percent of staff completing the non-clinician survey around the start of the third program year reported that they strongly or somewhat agreed that the program had successfully engaged participants, while slightly more than 50 percent of staff reported that participant engagement was helpful to the program in

reaching its goals. About half of the staff completing the survey, however, also reported that the program may have required too much time and had too many requirements for the families.

c. Barriers and facilitators associated with service delivery effectiveness

Boston Medical Center's ability to effectively deliver intervention services was influenced by several factors. Looking over the full three-year cooperative agreement, staff honed in on two key facilitators of service delivery: (1) the multidisciplinary care teams composed of highly motivated staff and (2) staff use of ACT.md as a care management tool.

First, program leaders and staff described in interviews how the multidisciplinary teams, which had expertise related to the CMC's diverse medical, behavioral, and social needs, allowed team members to (1) triage tasks to those whose expertise was the best fit and (2) collectively meet the families' complex and interrelated needs. In addition, program leaders said they hired staff who were self-motivated and committed to the program and the population served.

Second, over the course of the cooperative agreement, 4C staff came to rely on the care management features of ACT.md to aid them in their work. In interviews, nearly all staff described ACT.md as a key resource for information sharing among the team members. A nurse care coordinator, for example, said that ACT.md was a critical tool in assigning tasks among various team members and in organizing and tracking their coordination activities.

Program leaders and staff also described four primary barriers to the delivery of intervention services: (1) difficulty maintaining the scope of services with increased caseloads, (2) insufficient staff training and protocols, (3) limited outside provider engagement, (4) challenges with patient and family engagement.

First, in both the second and third program years, some staff reported difficulty providing the same level of care coordination as in the first program year. Because caseloads were smaller than expected in the first program year, staff had more time to provide comprehensive, hands-on support to each CMC and family. As caseloads grew, staff had less time to conduct home visits or attend in-person appointments. This was particularly problematic for staff in Boston because they enrolled a high percentage of children with complex behavioral health and other psychosocial needs, which staff found to be more time-consuming to support than complex medical challenges. In addition, one nurse care coordinator left the program in the third year. Program leaders decided not to fill the vacant position due to the uncertainty of program funding after the cooperative agreement, leaving one nurse care coordinator with a double caseload.

> "I think we maybe sold ourselves as a team that could do everything because we had a smaller caseload [in the first program year]. Then, when it picked up, we realized … first, we can't keep doing what we're doing because the caseload's larger and, second, what should the family be taking on for themselves or what [will they] need to be doing when we're not here anymore, and what are their ultimate goals for independence?"
>
> —Family navigator

When caseloads became unmanageable, staff said that they were often able to rely on other team members for support. Staff also learned that building strong relationships with and referring participants to external stakeholders (such as the Massachusetts's Department of Public Health and Department of Developmental Services) and community service providers (such as schools

and housing authorities) helped them to manage their caseloads. In addition, staff reported that efforts near the end of the cooperative agreement to disenroll less active participants and transition them to other community supports or programs within their medical systems gave the staff more time to focus on their active participants.

Second, some 4C staff described insufficient formal training and a lack of written protocols guiding their work as a barrier to service delivery. Although some staff felt that the lack of rigid training and protocols gave them flexibility in how to best meet the needs of the participants, most agreed that it created stress for those staff who desired more structure and guidance. A few staff reported struggling with the lack of direction, formal supervision, and insufficient training and protocols, especially at the beginning of the cooperative agreement when staff were trying to delineate roles and establish care coordination processes. Over time, staff learned to overcome this challenge by leveraging informal support from 4C colleagues and other staff in their respective medical systems. The social workers and family navigators at the Boston site, for example, mentioned accessing supports and training through non-4C teams that provided similar services in the medical system.

> "I was given freedom and a lack of guidelines…. I struggled with it, but at the same time acknowledged that because of this freedom I've probably done things and affected patients in a profound way and that if I was tied to policies and rigorous structure, I might not have been able to [do so]."
>
> —Nurse care coordinator

Third, the limited engagement in the program of PCPs and other providers made it challenging for staff to coordinate participants' care. Staff said that some PCPs were more interested in passing off CMC to the 4C program and letting the 4C program manage their care than they were in actively partnering with staff to coordinate that care. Other providers, they said, were simply difficult to reach, likely due to competing priorities. Staff also described how providers outside the program did not utilize the ACT.md platform as a communication tool or access or update information in the CMC care plans at the expected level. Staff cited providers' reticence to learn and manage an additional data system and their competing priorities as the main factors that limited use of ACT.md. Staff were partly able to overcome these challenges by being persistent in their communication efforts, searching in the EMRs for needed information, and reverting to phone calls and emails (rather than ACT.md) for communication.

Fourth, 4C leaders and staff described ongoing challenges with family engagement. They said many parents were overwhelmed by the enormity of their children and family's needs, which often included medical as well as behavioral, psychosocial, and socioeconomic challenges. These challenges were a specific target of the program, but also contributed to some families' inability to engage fully in program services and caused other families to rely too heavily on the teams for assistance. Interviewees said that fostering relationships with and understanding the participants' needs—and focusing on obtainable goals—helped them to overcome these challenges with some families. To lessen families' reliance on the program, staff increased their emphasis on building families' self-sufficiency in the third program year. Staff modeled processes for families to

> "We do this full psychosocial assessment and usually come out with 20 things that need to be worked on with the family. That's not only overwhelming for us as a team, it's overwhelming for the family…. I think we are doing a better job of helping families pick a select number of goals that are most important to them and really trying to hone them in on those."
>
> —Nurse care coordinator

make them less overwhelming, followed up with the families to verify that they were successful, and stepped in only when necessary.

Staff also described how families did not use ACT.md or leverage the electronic care plans as fully as anticipated. Despite having received tutorials on how to use the system, many families lacked the technological literacy to use the platform effectively or were too overwhelmed with other competing demands to prioritize learning how to use the system. In addition, ACT.md and the care plan were available only in English, which made the system more challenging for some families to use. Although 4C staff actively encouraged families to use ACT.md earlier in the cooperative agreement, they stopped doing so in the third program year. Staff instead relied on telephone calls and face-to-face contact as their primary modes of communication with families. They also helped families print copies of care plans to have on hand if the child went to the ED or met with a new provider.

> "[ACT.md] is not super approachable for a lot of the families that we're working with. [They] are just trying to survive day to day versus trying to figure out how to utilize this technology."
>
> —Nurse care coordinator

C. Assessment of perceived program effects on the delivery of care and outcomes

The majority of staff who responded to the non-clinician survey around the start of the third program year believed that the 4C program had positive impacts on the delivery of care and CMC and family outcomes (Table II.2). In addition, all staff who completed the survey strongly (65 percent) or somewhat (35 percent) agreed that the program was making a difference in meeting critical needs in the community; most indicated that the 4C program was very (35 percent) or somewhat (45 percent) effective in achieving its goals.

Table II.2. Staff perceptions of 4C program effects on care

The 4C program had a positive impact on:	Percentage of staff who agreed with the statement (N = 20)
The quality of care and services you provide to participants	100
Your ability to respond in a timely way to participant needs	95
The efficiency of care services provided to participants	94
Your ability to provide care or services that are responsive to participant preferences, needs, and values	95
Care services that are provided fairly to all participants	95
Access to care or services for all participants	95
Achievement of participants' health goals	95
Participant satisfaction	100
Participant quality of life	100
Care coordination	94

Source: HCIA R2 evaluation survey of awardee's non-clinician staff, July to October 2016

Note: The survey had 20 respondents overall. A small number of responses (one or two) were missing for some survey items.

In interviews during the third program year, program leaders and staff conveyed similarly positive perceptions of the 4C program's impact on care delivery and CMC and family outcomes. Program leaders and staff felt strongly that the 4C program offered valuable services to CMC and their families, particularly for the Baystate site, where the services were unique to their community.

Program leaders and staff also believed that the comprehensive and intensive services they provided likely decreased parental stress, reduced service duplication, improved CMC quality of life, and decreased hospitalizations. Their perceptions of the program's potential to decrease health care costs, however, varied. In reports to CMS and in interviews, program leaders and complex care pediatricians cited results from preliminary internal analyses of cost data that showed decreases in health care costs for CMC when comparing the six months prior to 4C participation to the six months after initiating participation, but then increases in costs after one year. Some 4C care team members suggested that reductions in short-term costs were likely to be achieved through decreases in service duplication as a result of care coordination services or decreases in ED use as CMC caregivers learned strategies for avoiding emergency care. However, other 4C care team members hypothesized that acute care utilization and costs would increase in the months following program enrollment as 4C teams worked to meet intense or previously unmet needs of participants. One complex care pediatrician posited that health care costs would remain high for some participants regardless of 4C services because they have chronic medical complications, while costs for other participants may stabilize after initial receipt of previously unmet services and supports.

> "When a family first gets to [the 4C program], I think there'd be a spike in costs because these families were not making their appointments or accessing services beforehand. Then, they start going to appointments that before they did not know they had. They start filling needed prescriptions and going to the neurologist and getting procedures done. They also now may have an in-home team. This is all costly in the beginning, but could prevent a hospitalization."
>
> —4C nurse care coordinator

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

Program leaders and staff perceived the 4C program to have a positive impact for participants. However, some implementation issues have implications for the interpretation and analysis of program impacts.

> "Getting [families] hooked in [to the 4C program] and getting a few negative stressors reversed and getting a little wind behind their backs can really transform the trajectory of families from a high-cost social disaster direction to [a] stable, engaged, and steady state."
>
> —Complex care pediatrician

First, the program's struggles to fully engage PCPs and specialists in CMC care coordination activities may have limited the program's ability to achieve desired outcomes because PCPs may have continued to order duplicative or potentially unnecessary services for the children or not capitalized on the program's coordination efforts or the 4C complex care pediatrician's expertise.

Second, total health care costs may decrease in the long term for some participants, but potentially not during the three-year cooperative agreement. Program leaders and staff reported that some children are so medically complex that they may require ongoing intensive support and may continue to experience episodic needs that require acute or emergency care. That said, the intensive and early intervention care coordination supports provided by the 4C program may prevent some complications, decrease the frequency or length of hospital stays, and lower the long-term risk of some children's placement in more expensive out-of-home settings, thus decreasing lifelong health care costs.

Third, although the awardee enrolled 365 participants by the end of the three-year cooperative agreement, only 69 percent of the enrollees, or 252 participants, were covered by Medicaid or CHIP. Therefore, the sample size is small. This may make it difficult to statistically detect impacts on outcomes.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Boston Medical Center's 4C program.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: Boston Medical Center

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018	152 fee-for-service (possibly 252 total including managed care)[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	3,491
Likelihood of all-cause hospitalizations	827
MDE sample size requirement to detect 20% effect	
Total expenditures	873
Likelihood of all-cause hospitalizations	207
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group and patient self-selection high/high refusal rate
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	Too early to determine due to delay in Medicaid data
Do claims identify the primary expected effects?	Some effects observed in claims data but important effects likely missing
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Too few treatment beneficiaries for primary outcomes
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	BMC is collecting beneficiary surveys, but local evaluator is doing analysis; exploring using implementation data on services provided and time spent on each activity

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

At this point, we do not anticipate being able to conduct a rigorous impact analysis for the awardee due to the low number of participating beneficiaries. (As shown in Table III.1, there were only 252 participating Medicaid beneficiaries when enrollment ended in August 2017.) The local evaluator is administering beneficiary surveys and plans to analyze the survey results. We plan to analyze results from our staff and beneficiary surveys. We will also explore the possibility of analyzing data that the awardee collected on its interactions with participants, such as in-person visits, electronic communications, and so on.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

The Boston and Baystate sites were actively participating in discussions with hospitals, payers, and other stakeholders as part of a recent statewide effort to implement regional Medicaid ACOs. Both sites were expected to reach agreements in preparation for launch of the ACOs in March 2018. One of the new ACOs, the Boston Accountable Care Organization with Boston Medical Center Health Net Plan, will include Boston Medical Center, its physician practices, most of its licensed and affiliated community health centers, and other providers. Baystate and its community health centers will join a managed care organization to form the Baystate Health Care Alliance with Health New England ACO.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who were most familiar with the payment model.

B. Description of the payment model

The program leaders adjusted their payment model development strategy in response to a recent statewide effort to implement regional Medicaid ACOs. A total of 18 health care organizations in Massachusetts will participate in MassHealth's ACO program beginning on January 20, 2018, as part of the state's five-year innovative 1115 Medicaid waiver. ACO formation was still in the early stages, but both the Boston and Baystate sites were actively participating in discussions with hospitals, payers, and other stakeholders. They expected to reach agreements with their respective ACOs in preparation for launch in March 2018. One of the new ACOs, the Boston Accountable Care Organization with Boston Medical Center Health Net Plan, will include Boston Medical Center, its physician practices, most of its licensed and affiliated community health centers, and other providers. Baystate and its community health centers will join with a managed care organization to form the Baystate Health Care Alliance with Health New England ACO.

Program leaders partnered with a health economist to analyze claims data to evaluate program costs and savings and inform negotiations about program services and staffing under the ACO model. Program leaders at both sites acknowledged that participation in their ACOs may require some changes in program eligibility and services and require reporting quality measures to monitor performance, but no additional details are available at this time.

C. Status of the payment model

Both the Boston and Baystate sites have begun preparing for their inclusion in the ACOs. Boston envisions a single team serving Boston Medical Center and multiple community health centers but will need to determine staffing needs if the team will also serve the broader ACO population when it is defined. Baystate is negotiating a per beneficiary per month (PBPM) fee with its ACO. Baystate anticipates a target PBPM of $100, which will include all program services except physician visits.

D. Factors associated with the development of the payment model

In addition to ongoing changes and uncertainty related to statewide Medicaid reform efforts, program leaders identified several other challenges they've encountered in developing a payment model for their pediatric complex care programs. First, billing systems do not capture care coordination work for the management of complex care pediatrics, which makes it difficult to estimate the costs of the program. Program leaders have used claims data to assess the ratio of costs for hospital and ED visits and costs for in-home services for program participants by using place of service codes. They compared the cost ratio pre-enrollment to the cost ratio post-enrollment as one approach to evaluate program costs. Second, the awardee reported that payers initially expressed excitement about working with program leaders, but the payers' legal teams raised several concerns and ultimately prevented data sharing. The legal concerns included risks associated with the transfer of patient health information and data sharing that could reveal differences in contracted rates to an organization that was also working with competitors. Lack of payer data prevented program leaders from identifying additional children with high utilization who could benefit from the intervention and affected their ability to meet enrollment targets. One awardee leader suggested that the awardee may have had more success if it had collaborated and contracted with one payer rather than trying to get as many payers involved as possible, thus alleviating payer concerns about sharing proprietary financial information.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Although the awardee was awaiting the launch of its payment model (Medicaid ACOs), program leaders at both sites (Boston and Baystate) worked to focus the 4C program on its most effective components to help sustain it after the cooperative agreement. The Boston site appeared to have internal support from its hospital system for sustaining a program, but its partner, Baystate, did not. Both sites were trying to identify funding and strategies for scaling the program to other providers who will be involved in the ACO, as well as the participants who will not be included in the ACO. There appeared to be little activity to replicate the program, although the awardee planned to use part of its no-cost extension to disseminate information about the program throughout the country.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2 of the award, Boston Medical Center had actively begun pursuing strategies for sustaining the 4C program. The awardee explored three main potential sources of future funding: (1) engaging Medicaid managed care organizations, (2) joining new regional ACOs, and (3) using encounter data to inform and develop a PBPM fee to support program services. At that

time, Boston Medical Center hoped to expand the program to more locations, increase the number of care teams at each site, and expand clinic hours to improve access to program services, but had not yet implemented any scaling activities. The awardee reported no plans to disseminate information and assistance to other organizations to facilitate program replication.

C. Implementing the SSR plan: progress and changes

Sustainability. Integrating the 4C program into the new regional Medicaid ACO model (as discussed in Chapter IV), was the awardee's key strategy for sustaining the program, although the ACO launch was not expected until March 2018. The Boston site received indications of support from leaders at its hospital system to sustain a modified version of the program in the near term, but the Baystate site did not expect to have sufficient resources to continue the program without payer support. The awardee's six-month no-cost extension provided more time to analyze data and demonstrate positive program impacts to help justify the payment model recommendations to state and commercial payers and obtain additional funding to sustain and expand the model in one or both locations.

The sites made some changes to the 4C program to make sustaining it more appealing to hospital leaders and payers. Many of the changes involved focusing on the most important and effective aspects of the 4C program and downsizing to control administrative costs:

- Boston Medical Center was targeting outreach to more specific cases (namely, complex mental health cases) and to the PCPs within its own hospital system. Program leaders were also considering revisiting the eligibility criteria to better align with the ACO's requirements.
- Both sites planned to wind down community outreach efforts and home visits conducted by social workers and rely more on family navigator outreach and support. They also planned to pivot to a "leaner" care plan, which would highlight and provide accountability for key next steps for the providers and participants' families. This new plan would replace a nuanced and complicated medical history requirement that was difficult to update and share.
- The Boston site was considering switching from an open-ended to a limited enrollment period. Because awardee leaders speculated that the program provided its most intensive services in the first six months, the Boston site was considering transitioning to a stabilization program. Program staff would set a goal for each patient to reach within three to six months, then reassess the patient's status and potentially renew enrollment if warranted. Program staff also might try to determine at enrollment which participants might have persistent needs due to a chronic medical complexity and instead put them under the care of complex care physicians, rather than under the 4C care team. The Boston site was starting to disenroll inactive participants and appeared to be embedding the care of long-term enrollees to a particular hospital department or care team. To prepare the families (especially those of children most recently enrolled) for these changes, program staff started describing the program as a short-term "booster" program to initiate care, but explained that the child would transition to a moderate care management program.
- Both sites were contemplating changing the program to align with the future ACOs in their regions. Throughout the course of the cooperative agreement, the program was intentionally kept separate from broader care coordination work. However, the Boston site needed to

determine how to support the program components in the future ACO's distal sites of care, potentially by regularly visiting the sites or by using telemedicine. It was also considering embedding or linking the four nurse practitioners that conducted moderate case management into the new population health department formed as part of the ACO. Both hospitals were considering discontinuing the ACT.md electronic care coordination tool. The Boston site was contemplating switching to the ACO's case management system, which was trying to integrate some of the ACT.md functions, mainly those that enable patient interactions.

Scalability. The Boston site was interested in scaling a potentially modified 4C program to other community health centers and other provider sites, but the specifics had not yet been worked out. Baystate was beginning to find ways to integrate the program with a community health worker effort. Both sites were attempting to work with the state to potentially create a statewide 4C program for non-Medicaid participants who remain outside of the ACO. Because the current program serves about 1,500 of the 7,500 top-utilizing children in the state, a Baystate respondent noted that there were opportunities to reach many more children. The non-ACO model would involve the Boston and Baystate hospital systems plus several other hospitals (for example, the University of Massachusetts) aligning with the other Title V programs, including the Massachusetts Child Psychiatry Access Project. To gain support for this expansion, the partners were considering applying for a Robert Wood Johnson Foundation health pioneering grant. In addition, the complex care pediatrician at the Boston site thought the 4C program potentially could expand to an adult population. Adult medicine providers reportedly saw opportunities to learn from the 4C program's team-based approaches, care plans, and care coordination strategies.

Replicability. There was no evidence that the 4C program had been replicated elsewhere and neither the Boston nor Baystate sites appeared to be promoting program replication. However, one of the stated goals of the awardee's no-cost extension was to increase market awareness throughout the country of the program's success.

D. Factors associated with progress toward implementing the SSR plan

The 4C program garnered significant support and attention from leaders at the Boston site, which should help secure its sustainability. The ACO initiative helped boost interest in the program because the ACO designated the special needs and complex care pediatric population a priority. In addition, the ACO viewed the program's attention to complex care, care management, and care coordination as key to treating this population. As one Boston site respondent said, "We are probably in the top three bullets at every conversation around pediatric ACO readiness."

However, Baystate was struggling to gain internal ongoing support from its hospital system for sustaining the program. Program staff facilitated a letter-writing campaign by clinicians and families of participants to urge leadership to continue the program. They argued that sustaining the program was "the right thing to do," even though program leaders lacked evidence of cost savings in the short term and they had not had enough time to study whether it saved money in the long term.

Changing the program in the ways described could create a barrier to SSR. One awardee leader at the Boston site reported concern that removing any element could alter the program's impact in ways not anticipated: "You can have a tripod and take one of the legs off, but which direction it's going to fall you never know."

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: CareChoice Cooperative

April 18, 2018

Julia Doherty, L&M Policy Research
Maya Jean-Baptiste, L&M Policy Research
Poonam Pardasaney, RTI International

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

This page has been left blank for double-sided copying.

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

CareChoice Cooperative, a cooperative of skilled nursing facilities (SNFs), senior independent housing, and assisted living communities in Minnesota, used HCIA R2 funds to

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Care Choice received a four-month extension through December 31, 2017, to allow additional time for the awardee to analyze the program's impact and pursue its payment model.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

pilot the Person-Centered Care Connections (PCCC) program. The program was designed to improve the care and safety of post-acute care patients in SNFs who were transitioning back to the community from short-stay or transitional care units (TCUs) within the participating SNFs, and to reduce the total cost of their care.[4] CareChoice expected a 20 percent reduction in hospital readmissions that took place within 30 days post-discharge with the accompanying expectation that the intervention would yield a 3.5 percent reduction in total Medicare spending for program participants.

The PCCC program offered an innovative approach to care transitions by ensuring a more comprehensive transition planning process: web-based decision support software called Engage[5] was used to guide staff through this process. The PCCC program had four distinct components: (1) transitional care coordination, (2) patient and family engagement, (3) quality improvement (QI) and workflow process redesign, and (4) education and training. It launched in 10 CareChoice SNFs on January 1, 2015, with a goal of serving 8,874 participants during its three-year cooperative agreement.[6]

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	SNFs hired transition coordinators who used a web-based decision support tool (Engage) to improve the process of transitioning participants back into the community. The goal was to prevent unnecessary hospital readmissions, lower the total cost of care, and improve patient and family satisfaction.
Major innovation	Used additional staffing (hiring of transition coordinators), decision support IT software (Engage), and process improvement to give safer, better quality care to participants who were being discharged to the community.
Program components	1. Transitional care coordination: Improving the transition process with more communication between members of a multidisciplinary transition team[7] and more comprehensive transition documentation for patients and caregivers. 2. Patient and family engagement: Improving education for participants and families about diagnoses and the transition plan during participants' TCU stays. 3. Quality improvement and workflow process redesign (QI): Leveraging a web-based decision support tool (Engage) and processes to improve transition planning services. 4. Education and training: Giving a multidisciplinary team more training about its role in process improvements to ensure participants smoothly transition back into the community.
Target population	TCU patients in participating SNFs who are returning to the community
Payment model	New fee-for-service (FFS) payment or value-based payments, to be determined

[4] A TCU is a unit within a SNF that serves patients admitted for a short period of time with the goals of rehabilitating them and transitioning them back into the community.

[5] The Engage tool was developed by a vendor, Align, which is an organization focused on improving care transitions in post-acute care settings.

[6] Although not all TCU patients were included in the program's final evaluation data, all of them, regardless of payer, received transition services associated with the PCCC program and are referred to as participants in this report.

[7] The multidisciplinary care team was composed of staff working on the TCU floor and related departments, and included unit nurses, physical therapists, dieticians, and social workers.

Table III.2 *(continued)*

Program characteristic	Description
Award amount	$3,347,584
Effective launch date	January 1, 2015
Program setting	TCUs in participating SNFs
Market area	Urban, suburban
Market location	MN
Target outcomes	• Reduced total cost of care • Reduced hospital readmissions • Increased patient and family satisfaction • Increased patient and family understanding of the discharge plan

TCU = transitional care unit; SNF = skilled nursing facility.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on four factors. First, the awardee enrolled 8,016 PCCC participants—90 percent of its enrollment target—by the end of the initial cooperative agreement. Second, the awardee successfully hired program staff, introduced the Engage software, and completed the associated process improvement activities and training on schedule. Third, CareChoice worked with staff at each facility to closely monitor implementation and performance through data collection and analysis. Finally, program and SNF staff reported that the program had a positive effect on the delivery and quality of care, with facilities actively engaging participants in transition planning by sending them home with a comprehensive transition plan. Post-discharge surveys administered by program staff revealed that participants gave high ratings to the PCCC program's care delivery throughout the cooperative agreement.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of CareChoice's PCCC program, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants and the results of our efforts to identify a matched comparison group.

Payment model. CareChoice proposed a fee-for-service (FFS) reimbursement model for enhanced discharge planning. Billing codes for these services currently do not exist. The awardee is continuing analyses during the no-cost extension period and did not anticipate implementing its payment model before the end of the cooperative agreement.

Sustainability plans. Without a payment model in place at the end of the cooperative agreement, CareChoice did not plan to sustain the PCCC program as a whole. However, its 10 implementing sites are continuing many aspects of the program in different ways. They are often embedding PCCC services provided by existing staff within their normal care delivery process. The PCCC program had not been scaled or replicated, and the awardee was more focused on developing its payment model during its no-cost extension period.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source was a SNF staff survey on staff members' perceptions of the program's effect on the delivery of care. This survey was fielded from July to October of 2016 with 82 staff members and achieved a response rate of 78 percent. The survey included nurses, social workers, nursing assistants, dieticians, physical therapists, and transition coordinators. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that CareChoice was successful in implementing its PCCC program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

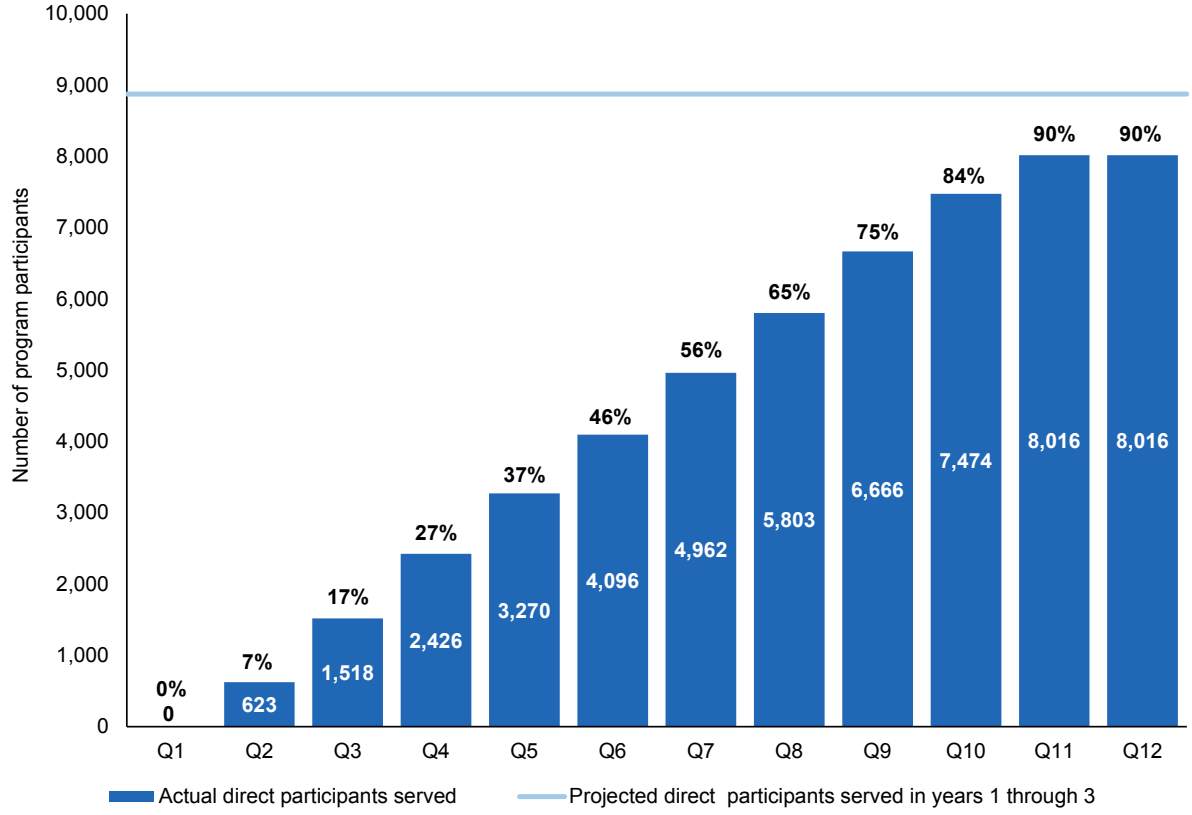
The PCCC program did not identify, recruit, or actively enroll participants. All patients admitted to a TCU in a participating SNF were automatically enrolled in the PCCC program and followed for 90 days post-discharge. Patients in SNF long-term care units were not enrolled in the program. Program participants from TCUs were considered disenrolled from PCCC if they were discharged to hospice, long-term care, or another TCU or hospital or if they left against medical advice. PCCC's passive enrollment process meant that TCU occupancy rates were the primary factor influencing whether or not CareChoice achieved its enrollment goals.

Before CareChoice launched PCCC, it planned to enroll 12,060 total participants during the three-year cooperative agreement. However, they reduced this target to 8,874 in Year 1 after one SNF dropped out of the program because planned renovations to its TCU were delayed. Additional enrollment challenges across SNFs are described in Section B.1.c below.

b. Evidence of enrollment effectiveness

Overall, the awardee reported that it enrolled 8,016 participants from January 2015 (when it launched its program) through August 2017, which represents about 90 percent of its final revised three-year projection (Figure II.1). At the beginning of the cooperative agreement, CareChoice planned to enroll 12,060 total participants during the three-year award. The awardee reduced this target to 8,874 participants in Year 1 after one SNF dropped out of the program. One reason that CareChoice gave for falling short of its revised projection is explained in detail in the next section on enrollment barriers and facilitators.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. At the beginning of the cooperative agreement, CareChoice planned to enroll 12,060 total participants during the three-year award. The awardee reduced this target to 8,874 participants in Year 1 after one SNF dropped out of the program.

c. Barriers and facilitators associated with enrollment effectiveness

> "One thing is that we have a preferred partnership with [a] hospital system, and we had to be approved, but I think because we are a part of the grant, it allowed us to form this partnership. It happened a year and a half ago. The application process was six months long. It was really intense. But we are really proud of that."
>
> —Transition coordinator

The awardee's progress in meeting its enrollment goals was primarily influenced by the participating SNFs' TCU occupancy rates. Occupancy rates, depending on whether they were above or below previous rates for a given SNF, functioned as either facilitators of or barriers to enrollment. Some SNF staff noted market pressure was the primary cause of the lower occupancy rates. For example, one participating SNF reported having a lower rate in Year 3 because a new SNF opened nearby and offered private rooms. Local payers and accountable care organizations (ACOs) also had an impact on participating SNFs' ability to attract new patients. The awardee reported that many payers and local health systems were narrowing their networks of SNF providers and did not select every PCCC SNF to be part of their preferred networks. Some SNF staff reported that being part of PCCC increased their chances of being selected, however.

2. Delivery of program services

a. Description of service delivery model

All CareChoice facilities that participated in the PCCC program were part of an earlier CareChoice initiative called Resident Centered Care Connections (RCCC), which was designed to help facilities reduce avoidable hospital admissions using core principles of Project Re-Engineer Discharge (Project RED[8]). Project RED principles were also integrated into the PCCC service delivery model to improve transitions specifically for patients returning to the community following a stay on a TCU.

Although SNFs were already implementing many of the activities that were part of the awardee's program to some degree even before HCIA R2, PCCC was a more comprehensive and systematic approach to care transition planning. Although the Engage software[9] existed prior to PCCC, SNF staff were not using it either as part of RCCC or Project RED. As part of the intervention, staff were encouraged to use the Engage software and the newly hired transition coordinators to (1) teach participants and family members about care management through short educational modules, (2) monitor the steps required to develop comprehensive discharge plans, and (3) produce a comprehensive package describing post-discharge instructions and services. Participating SNFs used HCIA R2 funding to buy the Engage software and hire one full-time

[8] Project RED (Re-Engineered Discharge) is a nationally recognized guiding framework for improving discharge planning and the transition process. See https://archive.ahrq.gov/professionals/systems/hospital/red/module1/index.html.

[9] The Engage software is a web-based decision-support tool developed by Align. The tool is intended to be used by all staff involved in the transition planning process—unit nurses, physical therapists, dieticians, and social workers—throughout a resident's SNF stay.

transition coordinator at each facility. Transition coordinators led transition planning activities and used Engage to collaborate with the multidisciplinary care team.

Program leaders used the PCCC's key QI and education/training components to guide the multidisciplinary care teams as they worked on the other two key program components: transitional care coordination and patient and family engagement. These QI, education, and training efforts involved having PCCC staff review facilities' self-monitoring reports and action plans, PCCC and Engage monitoring reports, root cause analyses, and the results of program staff site visits during monthly meetings with the transition coordinators.

Transitional care coordination and education of participants and families began soon after participants were admitted. A nurse on the multidisciplinary team performed a medication reconciliation and a team member administered a patient experience survey by using the Engage tool. This allowed staff to determine the right medication regimen for the participants and to identify and address participants' needs early in their stay. Team members used lessons in the educational modules in the Engage tool to educate participants and their families about managing their diagnoses once they transitioned home. A lesson was considered complete once a member of the team reviewed the lesson in person with the participant or caregiver and gauged their comprehension, either by asking questions at the end of the lesson or evaluating the participant's level of engagement during the conversation.

With the support of the team, the transition coordinator began preparing a comprehensive transition plan upon a patient's admission to the TCU. Transition coordinators or other team members conducted a second patient survey shortly before each participant was discharged to assess the participant's readiness to make the transition. Staff then provided more education or support if a participant expressed concern about his or her readiness to transition back to the community. Staff also made an appointment with each participant's primary care provider (PCP) before discharge. A team member sent a detailed written transition plan home with the participant and/or the caregiver. It included a medication list and schedule, a description of the services provided, post-discharge instructions, a contact number for questions, and summaries of key learning lessons. A team member also sent the same plan to the participant's PCP and—beginning in the third year of the cooperative agreement—to the participant's pharmacist. Finally, a team member, typically a transition coordinator, made follow-up calls to participants 2, 30, and 90 days after the post-acute stay to address any concerns or issues following discharge.

> "Several times [a follow-up call] has allowed the patient to tell me that their home care agency has not gotten in touch with them, or that they went to the pharmacy to pick up medications, and one of them was missing. It allowed me to go back to those providers and rectify those problems and find out why they were experiencing those issues.
>
> —Transition coordinator

b. Evidence of service delivery effectiveness

Overall, CareChoice successfully achieved service delivery effectiveness. The awardee launched the PCCC program on time and demonstrated early signs of success by meeting many of its self-monitoring measure goals in the first program year, while showing somewhat consistent and continued improvement throughout the three-year cooperative agreement.

CareChoice and participating SNFs hired the necessary staff to implement the PCCC program. Program leaders experienced some challenges in meeting all the training needs and expectations of program staff, but they nonetheless engaged SNF administration, transition coordinators, and the multidisciplinary teams to implement services as intended. During the third year of the cooperative agreement, program leaders also involved medical directors of participating SNFs in monitoring PCCC program implementation. Overall, program leaders leveraged QI activities to improve service delivery, staffing and training, and provider and staff engagement.

We used a variety of data sources to determine the awardee's effectiveness, including self-monitoring measures reported by the awardee and qualitative data from site visit interviews and the SNF staff survey. Details of our assessment follow.

Delivery of intervention services. CareChoice reported successfully delivering intervention services as intended. Across TCUs, the awardee monitored between 18 and 35 measures, 7 of which reflected performance on implementing a systematic approach to transition planning. CareChoice exceeded its projected targets on each metric during the cooperative agreement, and achieved nearly 100 percent of its targets for these metrics by the beginning of the third year.

Participating SNFs had relevant experience from implementing the earlier RCCC program, and facilities generally achieved the targets established by CareChoice from the beginning of the cooperative agreement. For example, during the first quarter of Year 1, participating facilities sent 95 percent of participants home with a written transition plan, and sent transition records to PCPs on time 87 percent of the time.

Although following up with patients over the phone after discharge was a relatively new practice for most participating SNFs, program staff met all related targets for phone follow-up almost every month of the three-year cooperative agreement. During the final year, in the aggregate, 91 to 100 percent of participating SNFs attempted to follow up with patients by phone 48 hours post-discharge, and 98 to 100 percent had attempted to follow up with patients by phone 30 days post-discharge.

The success rate of participating SNFs on verifying before discharge that a follow-up appointment with a PCP had been made was much lower (55 percent in Year 1, and 60 percent in Year 2) until the beginning of Year 3. A month into the final year of the cooperative program, leaders changed PCCC's approach, having staff automatically make PCP follow-up appointments for participants before they were discharged instead of encouraging participants to make an appointment with their PCP themselves. If a participant wanted a different appointment time, he or she could change it themselves or ask staff to change the appointment. CareChoice reported that SNFs significantly improved performance on this measure, going from a range of 20–58 percent in the proportion of appointments scheduled during the first two years to 87–100 percent during the final year of the cooperative agreement.

PCCC program staff reported that once they consistently met or exceeded original goals, program leaders sometimes raised the goals to foster continuous improvement. By the third year of the program, PCCC achieved between 95 and 100 percent of almost all of its target program

metrics. PCCC staff reported they accomplished this by implementing QI processes and using the Engage tool while giving transition coordinators ongoing support and supervision.

Staffing and training. Overall, CareChoice successfully staffed the PCCC program. CareChoice hired two key program leaders before launching the program, and both of them stayed with the program for the length of the cooperative agreement. Each SNF hired and managed its own transition coordinators and program leaders. SNF staff reported developing plans to cover PCCC functions in the event of team members' absences. Although turnover was somewhat common, participating SNFs reported that it was not difficult to replace transition coordinators when needed. However, it was challenging to train new staff to use the Engage software upon being hired, or to familiarize staff who did not regularly working on the TCU with the PCCC program and the Engage tool.

CareChoice delivered or arranged for significant staff training. The awardee provided a variety of opportunities for learning and skill development throughout the cooperative agreement: individual consultations, emails, monthly TC meetings, and occasional site visits, in addition to formal training. Seventy-seven percent of staff survey respondents said they attended formal PCCC trainings such as workshops, webinars, conferences, and presentations by phone, web, or in-person. According to the same survey, 72 percent of respondents somewhat or strongly agreed that the training helped them improve their job performance on the program.

Nonetheless, some transition coordinators indicated they would have preferred more intense training and hands-on assistance in using the Engage tool at the beginning of program implementation or whenever they first started. Several transition coordinators also reported that, at times, training was not provided in the appropriate format or time frame. During the Year 3 site visit, several SNF staff said they would have liked more hands-on training when they initially began using Engage. Similarly, 27 percent of the survey respondents indicated they somewhat or strongly agreed that they wished they had been offered other training to meet their responsibilities with the program.

Recruitment and engagement of providers. A key tenet of the PCCC program was to engage the multidisciplinary team and transition coordinators at all participating SNFs. These staff regularly communicated and collaborated during the transition planning process to offer participants a more comprehensive transition plan. Program leaders considered the fostering of communication and collaboration among program staff and program champions as major facilitators during the third program year. The majority of SNFs said they intended to either keep the newly hired transition coordinator on staff or roll the coordinator's tasks and responsibilities into existing positions after the PCCC program ended, reflecting their commitment to the program's ongoing improvement of care transitions. Almost all participating SNFs also said they expected to continue monitoring care transitions from their TCUs into the community more systematically then they did before the introduction of PCCC.

> "We see value with the transition coordinator and would love to continue that if we can. I don't know if our staff could do the program without her. But it is a hope of ours to expand the role and be able to follow up for even longer."
>
> —Director of nursing

Engagement of program participants. CareChoice successfully engaged program participants during the three-year cooperative agreement. Fifty-seven percent of respondents said participant engagement in the program was the most helpful factor to achieving PCCC program goals. Sixty-three percent of transition coordinators and multidisciplinary team members who responded to the SNF staff survey said that participant resistance to the program was not a barrier for PCCC.

On the patient satisfaction surveys administered by multidisciplinary care team members, participants gave PCCC care and services high ratings, helping the awardee to meet most of its patient satisfaction measures throughout the cooperative agreement. During site visit interviews, multiple staff also reported a significant increase in patient satisfaction after they implemented the PCCC program, and several indicated that participants particularly appreciated the follow-up phone calls post-discharge.

c. Barriers and facilitators associated with service delivery effectiveness

CareChoice leaders and staff said there were three key barriers to implementing PCCC: (1) duplicative data entry, (2) staff turnover, and (3) the Engage software training.

Duplicative data entry. First, throughout the cooperative agreement, frontline staff had to re-enter program data because the Engage tool and the SNF electronic medical records (EMRs) were not interoperable. Re-entering medication data was the most time-consuming task. Also, SNF staff had to submit additional facility metrics to PCCC on a third document type so PCCC staff could monitor facility-level data on a monthly basis. PCCC leaders agreed that the lack of software interoperability was a barrier CareChoice could not overcome during the cooperative agreement.

Staff turnover. Second, participating SNFs experienced challenges associated with staff turnover. Several transition coordinators and key staff left in the first year of the program and newly hired or temporary team members weren't familiar with the Engage software. (For example, while some staff may work on a particular unit most of the time, they may cover the TCU when needed, and not be familiar with Engage). In the SNF staff survey, respondents identified turnover as one of the top two barriers for the program. Staff interviewed in both the first and second program year reported that adapting to staff changes in the multidisciplinary teams and facility administration positions was a challenge.

> "More in-depth training on how to enter information in Engage is hugely important. I think what I got was pretty basic, and none of the details were covered. The other part is knowing all of the capabilities of the Engage software. There are a lot of features not utilized simply because we weren't aware of them."
>
> —Transition coordinator

In order to ensure transition planning at times of frequent staff turnover in Year 1 of the cooperative agreement, program leaders met with SNF administrators or their multidisciplinary team managers (such as a nurse supervisor or director of nursing) and each developed and implemented a sustainability plan to ensure the transition coordinator position and key team roles were covered so that transition planning tasks were performed when there were staff vacancies. Program leaders and transition coordinators said these plans mitigated the problems usually associated with staff changes and that turnover among transition

coordinators was no longer a barrier to implementation by the beginning of Year 3 of the cooperative agreement.

Engage software training. Most staff participated in formal training opportunities and the majority of staff survey respondents found Engage training helpful. However, some transition coordinators and team members noted they wished they had received more intensive, hands-on training on the Engage tool when they began to use the tool. Some staff also reported that they would have liked additional training on its more advanced tracking capabilities. In contrast, program leaders considered the training and related supervision and support provided to be a major facilitator for the program. The reason that the program staff and SNF staff had somewhat different perspectives may have stemmed from their perceptions about informal training opportunities and engagement in process improvement activities. PCCC staff worked closely on an ongoing basis with transition coordinators in integrating QI into their transition planning process, and program staff reported this served as a sort of less formal training, contributing to achieving the majority of PCCC's performance goals.

Overall, program leaders identified far more facilitators than barriers to the delivery of program services, including, but not limited to (1) fostering their communication and engagement before program launch, (2) successfully engaging the multidisciplinary team and other key staff, (3) having program champions, and (4) implementing PCCC and other related QI initiatives.

Fostering communication and engagement to include leaders before program launch. First, CareChoice began fostering engagement before the program launched by selecting SNF administrators who were eager to become champions for improving transition services following their experience with CareChoice's RCCC initiative. These administrators recognized the benefits of the funds that would allow them to hire transition coordinators and the support they would receive to use the Engage tool. The awardee recognized the importance of regular communication with SNF staff and transition coordinators in particular. After providing initial training, PCCC staff held monthly meetings with transition coordinators to help them learn how to use program data to improve processes and meet program goals. In interviews, staff interviewed said they gained valuable skills in data collection and analysis through the PCCC program.

Multidisciplinary team and senior staff engagement. Second, CareChoice encouraged transition coordinators to engage multidisciplinary staff in transition planning and work within the Engage tool to further facilitate systematic transition planning throughout the cooperative agreement period. Most transition coordinators said that getting other staff members on board was a challenge at first, but said team involvement improved during the length of the cooperative agreement. In addition, PCCC staff also reported greater engagement from upper-level SNF staff later in the cooperative agreement. For instance, medical directors played a significant role in resolving issues with local hospitals who were prematurely sending patients to participating SNFs in the third year, and staff said this avoided a number of unnecessary readmissions.

Program champions. Third, program leaders considered the transition coordinators to be the lead champions of the program on the front lines, and attributed much of the program's

success to frontline staff gradually adapting program interventions to better meet the needs of their facility and patient population.

Several staff and senior executives talked about a change in expectations and culture after Year 1 in terms of coordinating transitions, and noted the importance of the transition coordinators becoming a trusted part of the team. Staff had many examples of team members who took the initiative and championed the additional transition work. For example, several transition coordinators started compiling their own library of learning lessons that weren't available in the Engage software, such as energy conservation and deconditioning education.[10] Program leaders then worked with the Engage vendor to upload additional learning lessons. SNF staff also said some internal processes were changed to address participant concerns.

PCCC and complementary QI initiatives. Finally, CareChoice also reported that throughout the cooperative agreement, the PCCC program and participating SNFs were committed to continuous QI initiatives and the use of data to determine the root cause of any deficiencies. Staff also said they benefited from complementary local and national QI initiatives that, alongside PCCC efforts, facilitated better workflow and more attention to care transitions. The SNFs' earlier participation in the RCCC initiative prepared facilities to take further steps to support participants so they would be able to avoid unnecessary readmissions. In the final year of the cooperative agreement, supporting multiple ongoing initiatives, PCCC staff encouraged transition coordinators and team members to (1) obtain a Physician Orders for Life-Sustaining Treatment (POLST) form,[11] (2) offer palliative care when appropriate, (3) increase focus on other potential contributing factors to avoidable conditions that cause readmissions, such as falls, decreased functional status, and sepsis, and (4) send updated medication lists to participants' pharmacists and their PCPs upon discharge.

C. Assessment of perceived program effects on delivery of care and outcomes

CareChoice staff and leaders thought the PCCC program had a positive effect on the transition planning and quality of care that were delivered. More than two-thirds of the respondents to the staff survey (68 percent) said PCCC made a difference in meeting critical needs in the community. Several staff had examples of times when they got participants to reconsider going to the hospital in non-critical situations by using the resources offered by the program, and reported that they were able to help participants access services or answer questions during the PCCC follow-up calls.

[10] Deconditioning is a complex process of physiological change following a period of inactivity, bed rest, or sedentary lifestyle, common in the elderly. It results in functional losses in such areas as mental status, degree of continence, and ability to accomplish activities of daily living (*Canadian Nurse*, vol. 101, no. 6, June 2005, pp. 16–20).

[11] POLST is a form used by providers to discuss patient's wishes for end-of-life care with them and document related medical orders to be honored by health care workers during a medical crisis.

Transition coordinators and multidisciplinary team members reported spending more time educating participants and their family members during the transition planning process because of the PCCC program. In the staff survey, 55 percent of respondents (and 75 percent of transition coordinator respondents) reported that their ability to provide care or services to patients was better than it was before the PCCC program launch date, and 59 percent of respondents said the same about providing care or services to families and/or caregivers. Program staff and leaders believed these higher levels of engagement among participants and their families or other caregivers, along with more consistent and comprehensive transition planning, likely had a positive effect by reducing hospital readmissions and Medicare spending.

> "A lot of times, folks wanted to go to the hospital because they had too much pain. [PCCC] allow[ed] us to have a discussion with them about what we could change about their current pain management plan, and try it out before sending them in [to the hospital]."
>
> —Transition coordinator

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

The multitude of local, state, and national nursing home QI campaigns introduced in the past decade[12] have made SNFs increasingly attentive to the need for continuous QI activities, and may make it difficult to differentiate between the impact of PCCC interventions and other ongoing quality initiatives. In addition to the growing number of initiatives, several payers and both Minneapolis hospitals are developing preferred nursing home networks, putting even more pressure on SNFs to demonstrate high quality performance and value in order to maintain their census.

CareChoice piloted the RCCC program, a precursor to the PCCC program, from mid-2010 to mid-2013. All 10 participating SNFs also participated in this pilot program. The existence of this program, which predated our baseline period (2014), may affect the ability to select a valid comparison group and reduce the measured effect of the CareChoice intervention. It is possible that PCCC participants were entering a facility whose mean outcomes were affected by the RCCC program.

[12] The PCCC program began soon after CMS introduced financial penalties for hospitals with high readmission rates. What is now known as the National Nursing Home Quality Improvement Campaign lists reducing hospital readmissions as one of its four key goals. The Quality Improvement Organization for Minnesota also focuses many of its efforts on improving nursing home quality and care transitions, and Minnesota has its own Nursing Home Quality Report Card, in addition to providing consumers data on the CMS five-star rating system.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of CareChoice's PCCC program, baseline characteristics of the treatment group, and the results of our efforts to identify a matched comparison group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: CareChoice

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	1,425[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	458
Likelihood of all-cause hospitalizations	763
MDE sample size requirement to detect 20% effect	
Total expenditures	115
Likelihood of all-cause hospitalizations	191
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented intervention building in part on some components of an earlier program
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	No serious issues; proceeding with comparison group selection
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff survey
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We have constructed a comparison group from Medicare beneficiaries who entered a SNF in any of 12 counties in Minnesota. The treatment and comparison beneficiaries match well on all variables, so we believe that this matched group will yield credible estimates of program effects.

B. Characteristics of Medicare and Medicaid participants at baseline

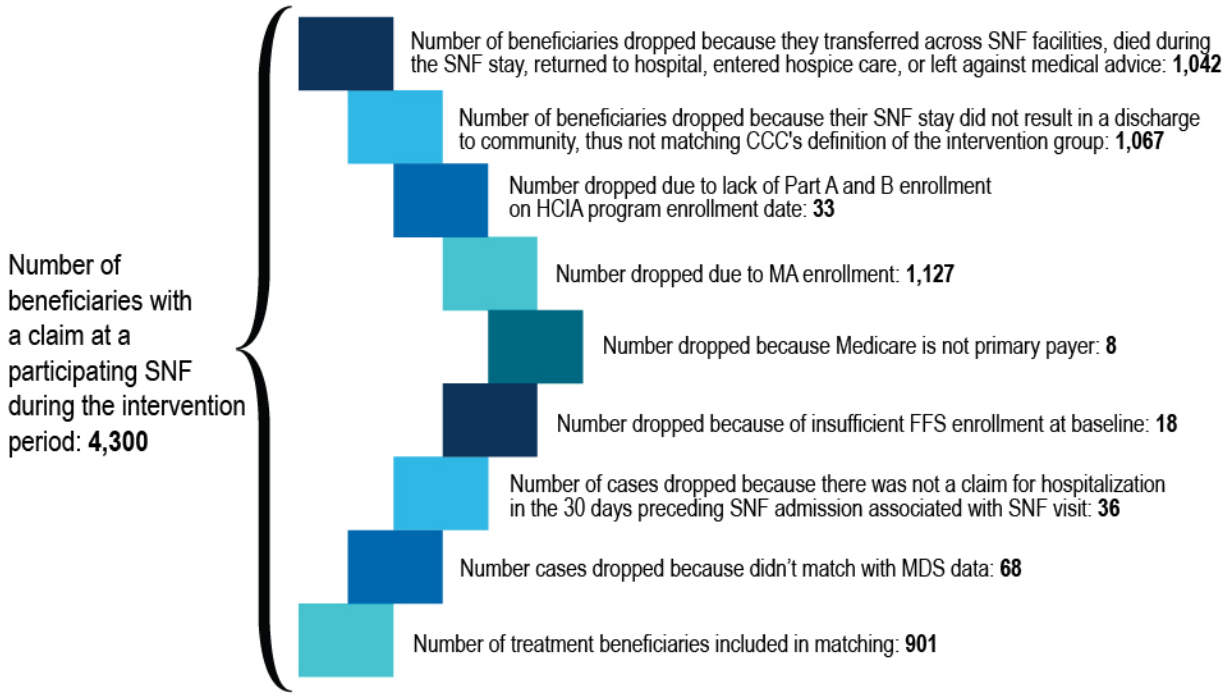
This section summarizes findings about the baseline characteristics of beneficiaries admitted to CareChoice facilities, as identified from Medicare SNF claims. To be eligible for the PCCC program, a prospective participant had to be a post-acute patient who was admitted to one of 10 participating CareChoice SNFs between January 1, 2015, and April 30, 2017. Only beneficiaries discharged home were included in the target population. The target population excluded participants who died during the SNF stay, were enrolled in hospice care, left against medical advice, returned to the hospital during the SNF stay, or transferred to long-term care or another SNF. Beneficiaries were considered to be enrolled in the program from the day of a SNF admission until 90 days after a SNF discharge home.

We did not rely on the finder file because no parallel file existed for any potential comparison group. To ensure the treatment and comparison groups were defined identically, we extracted SNF claims for beneficiaries who were admitted to participating SNFs from January 1, 2015, through April 30, 2017, which was the awardee's enrollment end date. We applied the awardee's inclusion criterion of discharge to home by using the SNF claims Patient Discharge Status Codes 01, 06, 81, and 86. This resulted in a sample of 2,191 beneficiaries.[13] The same codes will be used to select the comparison group.

For the purpose of presenting baseline characteristics in this report, we further restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS (both Parts A and B), with Medicare as the primary payer at the time when their eligibility for awardee-provided services began (that is, their enrollment date) and who had met all evaluation criteria for a period of at least 90 days within the baseline year (the 365 days immediately before their enrollment). We restricted the sample to stays that were preceded by a hospital discharge in the 30 days before SNF admission. Medicare rules require a hospital stay within the 30 days preceding SNF admission; once discharged from a SNF, patients can return to the SNF directly within 30 days of discharge. About 93 percent of treatment beneficiaries had a hospital discharge in the 30 days preceding SNF admission. We therefore required a hospital discharge within this time frame to maintain the homogeneity of our sample, exclude outlier cases, and allow us to use variables from this hospital stay for the propensity score matching model. This resulted in elimination of 1,290 beneficiaries from the sample of 2,191 beneficiaries. A total of 901 Medicare beneficiaries met the eligibility criteria and were included in the analysis.

[13] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from extracted SNF claims. The number of participants in these claims who were admitted to participating SNFs from January 1, 2015 through April 30, 2017 differ from the number of enrollees provided to the implementation and monitoring contractor for several reasons. First, the time periods differ. In addition, we restricted enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS with Medicare as the primary payer). The impact analyses are further limited as described above.

Figure III.1. Reasons for exclusion of program enrollees from impact analyses



The calendar period covered by the baseline quarters is based on the enrollment date, which is defined as the participant's SNF admission date and therefore varies by participant. Demographic and health status characteristics of these participants are shown in Table III.2. The majority of participants are age 75 or older: 34 percent are ages 75 to 84, and 39 percent are age 85 or older. Only 10 percent of participants are younger than 65. Most participants are white (92 percent) and two-thirds are female (67 percent).

Sixteen percent of the participants were originally enrolled in Medicare because of a disability; 0.33 percent were originally enrolled because of end-stage renal disease (ESRD). Twelve percent are dually eligible for Medicare and Medicaid, although in some cases Medicaid benefits may be restricted to the payment of co-insurance and deductibles. The average hierarchical condition category (HCC) risk score for participants is nearly 2.5 times higher than the average score for Medicare FFS beneficiaries nationwide (1.0). More than three-quarters of the participants have HCC risk scores higher than the national average, with the 25th percentile being 1.48 and the 75th percentile being 3.23. These data indicate that PCCC program participants have substantially poorer health status and greater needs for care than most Medicare FFS beneficiaries.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the CareChoice program through April 30, 2017

	All participants (n = 901)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	86	10
65 to 74	161	18
75 to 84	302	34
85 and older	352	39
Gender		
Female	603	67
Male	298	33
Race		
White	825	92
Black	55	6
American Indian, Alaska Native, Asian/Pacific Island American, or other	15	2
Hispanic	2	0.22
Original reason for Medicare eligibility		
Old age and survivor's insurance	750	83
Disability insurance benefits	148	16
End-stage renal disease (ESRD)[a]	3	0.33
Hospice[b]	3	0.33
Medicare/Medicaid dual status, percentage dual[b]	107	12
HCC score[c]		Statistic
Mean		2.48
25th percentile		1.48
Median		2.15
75th percentile		3.23

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of April 30, 2017.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as SNF admission date. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

The participants had high rates of service use and Medicare expenditures in the 365 days before enrollment, which is consistent with their recent hospitalizations and SNF admissions. In Table III.3, we report baseline utilization and expenditure data for a common set of measures, including the four CMMI core measures. CareChoice expected to reduce hospital readmissions and outpatient ED visits for up to 90 days after a SNF discharge, compared with baseline year rates. If reductions in post-discharge hospitalizations and ED visits are realized, then the program should result in a similar decline in post-discharge Medicare expenditures.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the CareChoice program through April 30, 2017

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	901	878	889	901	901
Average Medicare expenditures PBPM[a]					
Total	2,509 (74)	993 (82)	1,026 (88)	1,026 (78)	6,907 (209)
Acute inpatient	1,567 (50)	280 (38)	338 (55)	339 (47)	5,243 (158)
Inpatient other[b]	78 (17)	72 (30)	23 (16)	6 (6)	210 (57)
Outpatient[c]	224 (14)	189 (20)	209 (22)	191 (18)	306 (23)
Physician services	457 (15)	269 (16)	283 (19)	301 (17)	968 (24)
Home health	84 (7)	72 (11)	80 (12)	88 (11)	95 (12)
Skilled nursing facility	63 (10)	90 (21)	68 (21)	61 (17)	35 (12)
Hospice	11 (7)	6 (6)	8 (6)	14 (8)	15 (9)
Durable medical equipment	24 (5)	15 (2)	18 (2)	26 (7)	36 (12)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	1,446 (30)	388 (44)	389 (50)	452 (49)	4,497 (58)
Outpatient ED visits[e]	776 (57)	671 (82)	679 (79)	761 (71)	990 (85)
Primary care visits in any setting	11,402 (278)	7,636 (342)	7,792 (323)	8,602 (366)	21,367 (565)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Primary care visits in ambulatory settings	6,982 (201)	6,304 (265)	6,602 (256)	7,049 (276)	7,933 (256)
Specialist visits in any setting	12,493 (366)	8,814 (395)	8,887 (453)	9,260 (376)	22,748 (718)
Specialist visits in ambulatory settings	7,677 (268)	7,152 (316)	6,968 (306)	7,591 (307)	8,901 (328)
Acute hospital admissions[d]	1,446 (30)	388 (44)	389 (50)	452 (49)	4,497 (58)
Measures of any health care utilization					
Percentage with a hospital admission[d]	99 (<0.5)	9 (1)	8 (1)	10 (1)	99 (<0.5)
Percentage with an outpatient ED visit[e]	40 (2)	13 (1)	13 (1)	14 (1)	18 (1)
Percentage with a 30-day readmission among all discharges	14 (2)	8 (3)	13 (4)	11 (3)	23 (4)
Percentage of participants with a readmission among all participants	5 (1)	1 (<0.5)	1 (<0.5)	1 (<0.5)	2 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of April 30, 2017.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. Because a small number of beneficiaries were discharged from these excluded stays prior to SNF admission, the reported percent with a hospital stay in the 12 months prior to admission is 99 percent rather than 100 percent as would be expected. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

In general, most cost and utilization measures were highest in the fourth baseline quarter, as would be expected given the presence of a hospitalization before SNF admission. We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $2,509, with the average PBPM Medicare payment for inpatient services ($1,567) being the largest driver of total cost of care. Physician services were the second largest drivers of cost, but were much less expensive than inpatient services, with average PBPM expenditures at $457 in the baseline year.

The rate of acute care hospitalizations was 1,446 per 1,000 Medicare FFS beneficiaries per year during the baseline year. As expected, the rate of acute care hospitalizations per 1,000 beneficiaries per year was highest in the fourth baseline quarter (4,497), compared with the first through the third baseline quarters (when the number ranged from 388 to 452). A hospitalization in the 30 days before SNF admission was a condition of eligibility; 99 percent of all beneficiaries had at least one hospitalization during the fourth baseline quarter.

In the baseline year, 14 percent of all hospital discharges in the treatment group were followed by a readmission in the 30-day post-discharge window, slightly below the national average for Medicare beneficiaries (18 percent). At the beneficiary level in the baseline year, 5 percent of all Medicare beneficiaries had a hospitalization with a readmission in the 30-day post-discharge window. It is important to highlight that the percentage of beneficiaries with a hospitalization was highest during the baseline fourth quarter (99 percent, compared with 8 percent to 10 percent in the first through third baseline quarters) and that a large proportion of those hospital discharges occurred on the last day of the baseline year. Thus, 30-day readmissions associated with hospital discharges in the last 30 days of the baseline year are not reflected in the baseline readmission measures, because those readmissions would fall beyond the end of the baseline year.

The rate of ED visits that did not lead to a hospitalization was 776 per 1,000 beneficiaries per year in the baseline year, with an increase in the fourth baseline quarter compared with the first through third quarters. In the baseline year, the rate of primary care visits in any setting was 11,402 per 1,000 Medicare FFS beneficiaries per year, and the rate of specialty visits in any setting was 12,493 per 1,000 beneficiaries per year. As expected, the rate of primary care and specialty visits in any setting was considerably higher in the fourth baseline quarter compared with the first through third baseline quarters, whereas the rate of such visits in ambulatory settings was slightly higher in the fourth baseline quarter.

C. Quality of matched comparison group

Because the intervention was not implemented using an experimental design, we relied on a matched comparison group design to examine the impacts of the CareChoice PCCC program. The results presented in this section demonstrate the high level of covariate balance achieved on a range of beneficiary, facility, and area characteristics that are likely to affect participation in the intervention and outcomes.

The goal of matching is to mitigate nonrandom selection of individuals into the treatment group. This is done by constructing a matched comparison group that appears similar to the

treatment group on key observable characteristics that affect treatment status, or are expected to affect outcomes. For the CareChoice intervention, we included matching characteristics at the beneficiary, facility, and area level, with beneficiary characteristics including demographics, baseline clinical and health status characteristics, and utilization history. Limiting the comparison group to Medicare beneficiaries whose characteristics closely match observed characteristics of the treatment group is also expected to reduce unobserved differences between the two groups if those characteristics are correlated with matching variables.

1. Description and identification of the treatment group

Mathematica used Medicare enrollment and claims data to identify treatment group beneficiaries. The 10 treatment facilities that participated throughout the intervention period were included in the treatment group. The facility that dropped out had only 13 beneficiaries before we applied the eligibility criteria below, with the last admission being in June 2015; this facility was not included in the matching process and will not be included in the impact evaluation. The PCCC program was targeted to patients discharged home, thus the treatment group was restricted to patients discharged home from participating SNFs, and included all patients in these facilities with such discharges. (That is, there was no consent required of participants, nor any selection of these discharged patients into or out of the program.) We defined the program enrollment date to be the SNF admission date.

We applied four criteria to construct the Medicare treatment group for propensity score matching and outcome analysis. All beneficiaries in the treatment group met the following four conditions:

- Admitted to a participating CareChoice SNF during the intervention period (January 1, 2015, through April 30, 2017) and discharged home
- Discharged from a Medicare inpatient stay within the 30 days preceding SNF admission
- Enrolled in Medicare FFS (both Parts A and B), with Medicare as primary payer on their program enrollment date
- Enrolled in Medicare FFS (both Parts A and B) for a period of at least 90 days during the baseline year (the 365 days immediately before their enrollment date)

Medicare coverage rules require a hospital stay within the 30 days preceding SNF admission; once discharged from a SNF, patients can return to the SNF directly within 30 days of discharge. About 93 percent of treatment beneficiaries had a hospital discharge in the 30 days preceding SNF admission; thus, we required a hospital discharge within this time frame to maintain homogeneity of our sample, exclude outlier cases, and allow us to use variables from this hospital stay for the propensity score matching model.

We identified 901 Medicare beneficiaries who met the eligibility criteria for the PCCC program as of April 30, 2017.

2. Identifying a potential comparison group

We used a two-step approach to identify a potential comparison group consisting of Medicare FFS beneficiaries admitted to non-participating SNFs during the intervention period

and discharged home. In the first step, we identified a set of potential comparison facilities using the 2015 Nursing Home Compare Provider data file, which contains data on SNF characteristics, including name, address, CMS certification number, and quality ratings. A facility was eligible to be a comparison facility if it was (1) a Medicare- or Medicaid-certified SNF located in one of the following counties in Minnesota: Hennepin, Anoka, Ramsey, Sherburne, Rice, Carver, Dakota, Goodhue, Le Sueur, Steele, Washington, or Wright;[14] (2) not-for-profit or government-owned; and (3) not a hospital-based or swing bed facility. In addition to the five counties where treatment SNFs were located (Hennepin, Anoka, Ramsey, Sherburne, and Rice), we selected comparison SNFs from seven additional neighboring counties to strengthen our sample size and statistical power. This was particularly important because all eligible SNFs in Sherburne County were treatment SNFs. After discussions with the awardee, we determined that we did not need to exclude non-participating CareChoice SNFs because there was no spillover of the intervention to these facilities. This process resulted in 63 potential comparison facilities.

As noted earlier, CareChoice facilities participated in prior programs focusing on reducing avoidable readmissions and improving discharge planning. Comparison facilities may well have participated in similar programs. There is no practical means of controlling explicitly for such participation in drawing inferences about the impact of PCCC. Instead, we control in both matching and impact estimation for the possible outcomes of such programs. Characteristics of both beneficiaries and facilities are matched as closely as possible in order to focus inference specifically on the effects of PCCC.

In the second step, we used Medicare claims to identify beneficiaries admitted to potential comparison facilities between January 1, 2015, and April 30, 2017, and applied the same eligibility criteria used to select treatment beneficiaries. As we did with the treatment group, we assigned the SNF admission date as the enrollment date for potential comparison beneficiaries. We identified 5,574 beneficiaries who met the eligibility criteria; these beneficiaries formed the potential comparison group for propensity score matching.

3. The matching process

a. Variables for matching

Table III.4 lists the variables used in the propensity score matching model. When determining which variables to use in the propensity score model, we considered factors associated with both treatment status and key outcomes such as total expenditures and readmissions. We included several beneficiary characteristics, including demographic characteristics, original reason for Medicare entitlement, comorbidities in the baseline year, clinical characteristics from the most recent hospitalization in the 30 days preceding SNF admission, and clinical characteristics at the time of SNF admission. We also included baseline measures of beneficiary utilization and expenditure because these are strong predictors of future spending. At the facility-level, we included SNF size and overall five-star quality rating from Nursing Home Compare data. At the area-level, we included an indicator of SNF location in a metropolitan county.

[14] All intervention facilities are Medicare-Medicaid dually certified and not-for-profit, except for one with government ownership.

Table III.4. Variables used for propensity score matching model, CareChoice

Matching variable	Data source
Beneficiary demographics and eligibility	
Age	Enrollment database
Gender (Male)	Enrollment database
Married	MDS 3.0
Race (White)	Enrollment database
Dual Medicare-Medicaid eligibility	Enrollment database
Original reason for Medicare eligibility (disability, ESRD)	Enrollment database
Variables from the most recent hospital stay before SNF admission	
Length of the most recent hospital stay	Claims
Intensive care unit or coronary care unit use in the most recent hospital stay	Claims
Variables from the SNF stay	
Case Mix Index (CMI) score at SNF admission (10 or less, 11–30, 31–50, 51+)	MDS 3.0
RUG-IV at SNF admission (rehabilitation-ultra high, rehabilitation-very high, rehabilitation-high, rehabilitation-other, other)	MDS 3.0
Activities of daily living score at SNF admission	MDS 3.0
Baseline year health status and utilization	
Baseline HCC score[a]	Claims
HCCs in the baseline year (HCC2: Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock; HCC18: Diabetes with Chronic Complications; HCC48: Coagulation Defects and Other Specified Hematological Disorders; HCC58: Major Depressive, Bipolar, and Paranoid Disorders; HCC84: Cardio-Respiratory Failure and Shock; HCC85: Congestive Heart Failure; HCC96: Specified Heart Arrhythmias; HCC108: Vascular Disease; HCC111: Chronic Obstructive Pulmonary Disease; HCC135: Acute Renal Failure)[b]	Claims
Total Medicare expenditure per month in the baseline year[c]	Claims
Beneficiary had a hospital readmission in the baseline year	Claims
Number of acute hospital admissions (per 1,000 beneficiaries) in the baseline year[d]	Claims
Facility-level variables	
SNF size (number of Medicare FFS stays admitted during the intervention period and discharged home)	Claims
Overall five-star quality rating	NHC 2015
Area-level variables	
SNF in metropolitan area	USDA Economic Research Service

[a]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. For all other participants, we used the HCC institutional algorithm.

[b]HCCs were calculated using claims diagnosis codes 2 through 25 from the most recent hospital stay, and all diagnosis codes on preceding hospital claims in the baseline year. The 10 most frequent HCCs were included in the model.

Table III.4 *(continued)*

[c]Total Medicare expenditures for the baseline year were calculated from all claims for each participant with at least one eligible day during that year.

[d]The hospitalization measure includes acute care hospital admissions and excludes all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

MDS = Minimum Data Set; NHC = Nursing Home Compare; RUG = Resource Utilization Group; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category; USDA = United States Department of Agriculture.

b. Other adjustments

We applied a caliper to the enrollment-date variable to ensure that treated beneficiaries and their matched comparisons had enrolled within 90 days of each other. We also applied a caliper penalty to the propensity score to encourage the algorithm to match treated beneficiaries to comparisons with propensity scores within one standard deviation of their own.

4. Results

a. Covariate balance

Figure III.2 displays the standardized differences on matching variables between the treatment and matched comparison samples.[15] Standardized differences measure the difference in weighted means between the treatment group and the matched comparison group on the standard deviation scale; in our implementation, we used the standard deviation in the treatment group to ensure that pre- and post-matching diagnostics are on the same scale and to align with the evaluation's focus on impacts in the treatment group. Standardized differences between the treatment and matched comparison groups for all variables were less than 0.10, indicating excellent covariate balance, with some covariates having standardized differences close to 0.

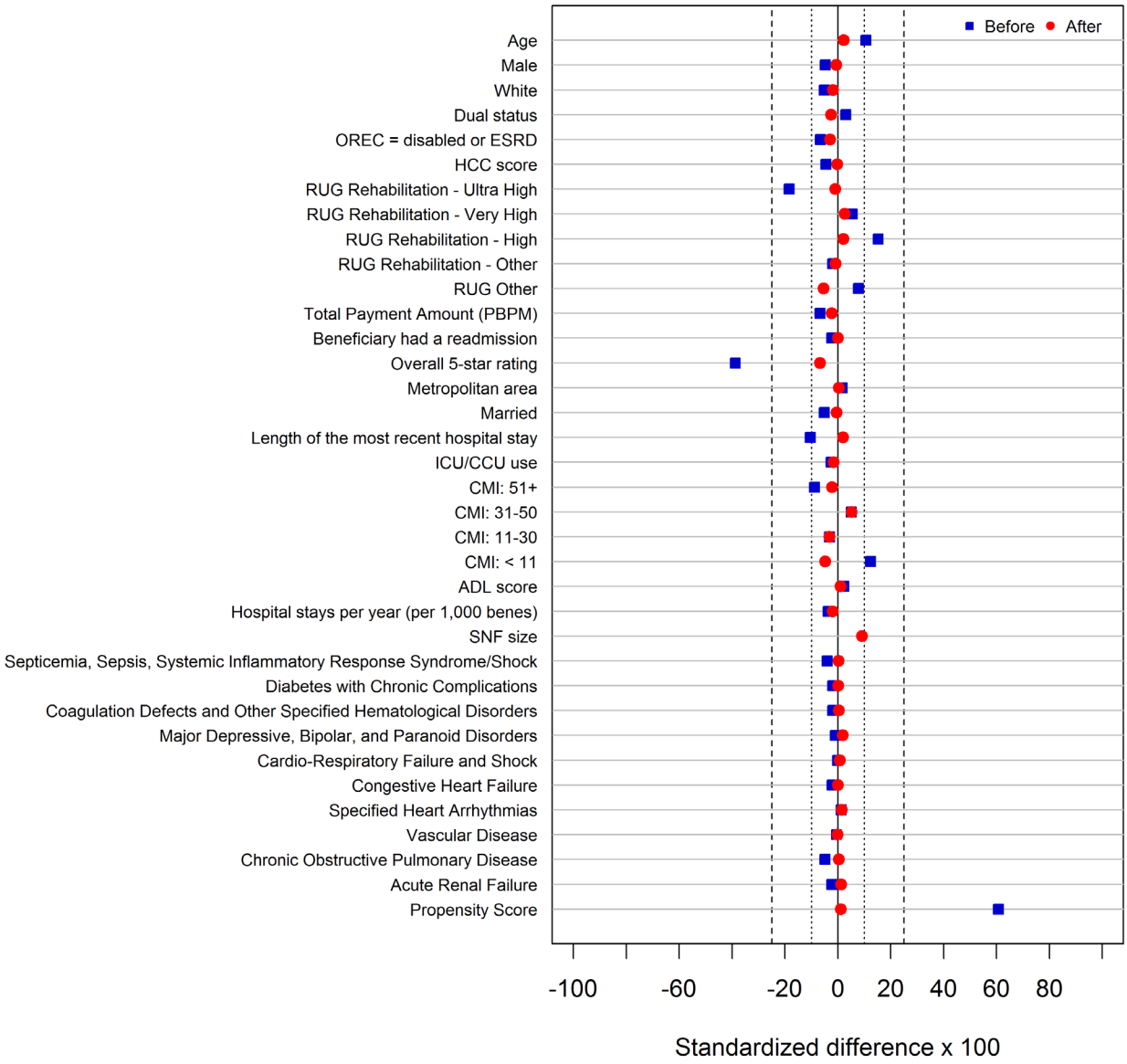
b. Final comparison group

The matching process yielded a total of 2,957 comparison beneficiaries from 63 comparison facilities, matched to the 901 treatment beneficiaries. Accounting for the matching weights gives an effective sample size of 1,688 comparisons. The distribution of treatment and comparison group beneficiaries by matching ratio (number of treatment beneficiaries matched to comparison beneficiaries in a set) is shown in Table III.5; most treated beneficiaries were matched to three or more comparisons.

Given the closeness of the means for treatment and comparison beneficiaries on all variables, we believe this matched group will yield credible estimates of program effects. Our regression analysis will control for the remaining differences between the two groups on these and other characteristics.

[15] A full balance table is included in Appendix A.

Figure III.2. Comparison of balance on matching variables before versus after matching for CareChoice



Notes: We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. For all other participants, we used the HCC institutional algorithm.

Total Medicare expenditures for the baseline year were calculated from all claims for each participant with at least one eligible day during that year.

The hospitalization measure includes acute care hospital admissions and excludes all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

OREC = Original Reason for Entitlement Code; ESRD = end stage renal disease; HCC = hierarchical condition category; RUG = Resource Utilization Groups; PBPM = per beneficiary per month; ICU = intensive care unit; CCU = coronary care unit; CMI = case mix index; ADL = activities of daily living score.

Table III.5. Distribution of treatment group beneficiaries by matching ratio, CareChoice

Matching ratio (treatment to comparison)	Number of treatment beneficiaries	Number of comparison beneficiaries
2:1	66	33
1:1	169	169
1:2	97	194
1:3	99	297
1:4	86	344
1:5	384	1,920
Total sample size	901	2,957

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Care Choice Cooperative has proposed an FFS reimbursement for enhanced discharge planning services. Billing codes currently do not exist for these services. The awardee is continuing analyses during its no-cost extension and did not expect to implement its payment model before the end of the cooperative agreement.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: (1) the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor and (2) narratives and payment model reports the awardee submitted to the implementation and monitoring contractor. We did not interview the awardee about its payment model because it did not advance from the development stage to active negotiations with one or more payers.

B. Description of the payment model

The CareChoice payment model was originally intended to be a one-time payment per patient discharge to cover the costs of enhanced discharge planning services to TCU/SNF patients; currently there are no quality metrics that would be used to adjust the payment. The awardee envisioned the payment would apply to each patient in a SNF and would cover the salary for the transition coordinator at that facility (one FTE), the cost of the software to guide the planning process, training, and any related supplies. The awardee envisioned that money saved by enhanced discharge planning—which would avoid unnecessary hospital readmissions or ED visits—could cover the costs of the intervention. The awardee believed that if the impact evaluation of the PCCC program demonstrated significant cost savings, this model could appeal to managed care organizations or to providers at risk of being responsible for their patients' expenses.

C. Status of the payment model

The awardee did not expect to have a payment model finalized and adopted by the end of the cooperative agreement for several reasons noted below (Section D). During the no-cost extension, CareChoice planned to conduct analyses that would help it develop a payment model for commercial payers. The awardee did not plan to sustain the intervention during its no-cost extension, although it was exploring whether there are codes similar to the Medicare chronic care management codes that could be used for this kind of intervention.

In one of the reports submitted to the implementation and monitoring contractor, the awardee also noted that "most of its facilities had opted to continue all or some of the PCCC interventions at their own cost as they believe that the positive impact and high level of patient and family satisfaction are worth the investment in staff, software, and resources needed to continue the project."

D. Factors associated with the development of the payment model

The awardee attributed the slow progress in developing its payment model to several factors. First, in its request to use unspent award money as carryover funds to support data analyses,

CareChoice noted that little of the initial award was dedicated to developing a payment model. Second, there were delays in obtaining claims data. Finally, the awardee reported that it wanted to conduct claims analysis for its payment model using an approach consistent with the evaluation team's, and chose to modify the analyses after talking with the evaluation team.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Without a payment model in place or near approval, CareChoice did not plan to sustain the PCCC program as a whole by the end of the initial three-year agreement. However, its 10 implementing sites continued many aspects of the program to varying degrees and in different ways, often by embedding the functions in their remaining staff, infrastructure, and functions. The program had not been scaled or replicated, and nascent activities to help foster such expansions stalled while the awardee focused on developing its payment model during its no-cost extension.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2 of the award, CareChoice had actively begun pursuing strategies for sustaining the PCCC program. The awardee reported that it had buy-in from site leaders and staff to continue the program. For example, CareChoice said leaders at implementing sites were strongly considering using internal funding to sustain their program staff and the Engage software. In addition, the awardee was developing a payment model by analyzing data acquired from two commercial payers and discussing the model with payers. The awardee was not working on

scaling the program at that time, but planned to scale the program in the future. The awardee was also developing a toolkit to help other entities replicate the program.

C. Implementing the SSR plan: progress and changes

Sustainability. As discussed in Chapter IV, without a finalized payment model (a case rate per discharge) CareChoice did not expect to be able to sustain the program as a whole immediately after the cooperative agreement ended, although the awardee hoped to pursue discussions about a model with commercial payers once it had collected enough data to show them. The awardee and its sites were taking different approaches to try and retain at least some parts of the program with existing resources and state grants. As noted, the awardee's four-month no-cost extension gave it more time to analyze the program's impact and pursue its payment model.

In a survey of its 10 sites at the end of the initial award period, the awardee reported that 2 sites are keeping the transition coordinator position, 8 sites will continue with the follow-up phone calls, and 6 out of the 10 will continue using the Engage software. Those sites that are discontinuing Engage or the transition coordinator position have chosen to try to incorporate key PCCC interventions into other people's jobs. Many of the sites reportedly will continue more comprehensive medication management, supported in part through a state grant that pays for pharmacists to regularly evaluate patients' medications and identify any unnecessary ones; staff other than transition coordinators will coordinate this process.

Sites vary in terms of how they plan to continue dedicating staff to sustain aspects of the PCCC program. Some sites expect their existing staff to be able to absorb the program functions, whereas others expect they will need to scale back staff (especially the transition coordinators) and some of the associated functions. For example, at one site, scheduling the appointments with PCPs before discharging a patient would reportedly be switched from a front desk function to a medical records function. At this same site, the interviewee noted the social workers will continue many of the functions that were performed by the transition coordinator, but in a less formal manner. Interviewees from multiple sites indicated that many of the PCCC functions have become "second nature" for staff so they expect they will continue, albeit with less reinforcement and oversight.

Scalability. CareChoice would like to scale the program to its other sites, and is educating and encouraging them to assess their transition processes. However, the awardee does not expect much scaling of the whole program because the other sites aren't necessarily using Engage, nor does the awardee yet have impact analyses that could convince these sites to implement the whole program.

Replicability. There are no signs of other organizations replicating the PCCC program. The awardee's planned development of the toolkit for others interested in implementing the program have stalled because the awardee needed to instead focus on improving the program and its impacts.

D. Factors associated with progress toward implementing the SSR plan

The awardee reported that many staff saw the benefits and positive impact of PCCC over the course of the program and have become champions for it. Most of the program has become embedded in many of the organizations. As one facility staff member put it: "This is the thing they breathe, this is their culture, this is what they do." PCCC leaders maintained that patients and family members will continue to benefit from the program after the award period, whether or not it receives ongoing funding. However, without an ongoing payment source, the program's sustainment will be inhibited, and scaling and replication are less likely.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS requests a follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

With matching now complete, we plan to begin developing the first round of impact estimates. These estimates will focus on standard outcomes such as hospitalization, readmission, ED use, and Medicare spending over two successive six-month periods after program enrollment. For each outcome, we will report the impact estimate, the 90-percent confidence interval for the estimate, and the *p*-value associated with the null hypothesis that the impact is zero. We will report these estimates to CMS as they become available.

This page has been left blank for double-sided copying.

APPENDIX A

POST-MATCHING DIAGNOSTICS

This page has been left blank for double-sided copying.

Table A.1. Post-matching diagnostics

Measure	Potential comparisons (n = 5574)	Matched comparisons (n = 2957)	Treated (N = 901)	Adjusted difference (matched)	Percentage difference (matched)	Standardized difference (matched)	*p*-value (matched)	Equivalence *p*-value (matched)
Age	78.686	79.585	79.820	0.234	0.300	0.022	0.588	0.000
	(0.148)	(0.200)	(0.359)	(0.504)				
Proportion male	0.353	0.334	0.331	-0.003	-0.800	-0.006	0.928	0.000
	(0.006)	(0.009)	(0.016)	(0.022)				
Proportion white	0.930	0.921	0.916	-0.005	-0.600	-0.019	0.715	0.000
	(0.003)	(0.005)	(0.009)	(0.013)				
Proportion dually eligible	0.109	0.127	0.119	-0.008	-7.100	-0.026	0.542	0.000
	(0.004)	(0.006)	(0.011)	(0.015)				
OREC = disabled or ESRD (proportion)	0.192	0.179	0.168	-0.011	-6.500	-0.029	0.526	0.000
	(0.005)	(0.007)	(0.012)	(0.018)				
HCC score	2.542	2.480	2.477	-0.004	-0.100	-0.002	0.976	0.000
	(0.021)	(0.028)	(0.048)	(0.069)				
RUG Rehabilitation - Ultra High (proportion)	0.524	0.437	0.432	-0.005	-1.200	-0.010	0.863	0.000
	(0.007)	(0.009)	(0.017)	(0.023)				
RUG Rehabilitation - Very High (proportion)	0.344	0.359	0.371	0.012	3.300	0.025	0.592	0.000
	(0.006)	(0.009)	(0.016)	(0.022)				
RUG Rehabilitation - High (proportion)	0.073	0.116	0.122	0.007	5.500	0.020	0.705	0.000
	(0.003)	(0.005)	(0.011)	(0.015)				
RUG Rehabilitation - Other (proportion)	0.019	0.018	0.017	-0.001	-7.000	-0.009	0.843	0.000
	(0.002)	(0.002)	(0.004)	(0.006)				
RUG Other (proportion)	0.040	0.070	0.058	-0.013	-22.000	-0.054	0.273	0.000
	(0.003)	(0.004)	(0.008)	(0.012)				
Total Payment Amount (PBPM)	2711.307	2614.786	2564.082	-50.704	-2.000	-0.023	0.634	0.000
	(35.393)	(49.323)	(73.410)	(116.957)				
Beneficiary had a readmission (proportion)	0.044	0.040	0.040	0.000	0.200	0.000	1.000	0.000
	(0.003)	(0.004)	(0.007)	(0.009)				
Overall 5-star rating	3.998	3.662	3.591	-0.071	-2.000	-0.067	0.134	0.000
	(0.012)	(0.018)	(0.035)	(0.050)				

Table A.1 *(continued)*

Measure	Potential comparisons (n = 5574)	Matched comparisons (n = 2957)	Treated (N = 901)	Adjusted difference (matched)	Percentage difference (matched)	Standardized difference (matched)	*p*-value (matched)	Equivalence *p*-value (matched)
Metropolitan area (proportion)	0.957	0.959	0.960	0.001	0.100	0.004	0.997	0.000
	(0.003)	(0.004)	(0.007)	(0.009)				
Married	0.354	0.332	0.330	-0.002	-0.700	-0.005	0.893	0.000
	(0.006)	(0.009)	(0.016)	(0.022)				
Length of the most recent hospital stay	6.502	6.032	6.107	0.075	1.200	0.020	0.649	0.000
	(0.056)	(0.062)	(0.128)	(0.167)				
ICU/CCU use (proportion)	0.244	0.240	0.233	-0.007	-3.000	-0.016	0.775	0.000
	(0.006)	(0.008)	(0.014)	(0.020)				
CMI: 51+ (proportion)	0.598	0.564	0.553	-0.011	-1.900	-0.022	0.726	0.000
	(0.007)	(0.009)	(0.017)	(0.024)				
CMI: 31-50 (proportion)	0.357	0.357	0.382	0.025	6.500	0.051	0.324	0.000
	(0.006)	(0.009)	(0.016)	(0.023)				
CMI: 11-30 (proportion)	0.029	0.029	0.024	-0.005	-20.300	-0.032	0.483	0.000
	(0.002)	(0.003)	(0.005)	(0.008)				
CMI: < 11 (proportion)	0.016	0.049	0.040	-0.009	-23.300	-0.048	0.369	0.000
	(0.002)	(0.003)	(0.007)	(0.010)				
ADL score	14.739	14.795	14.840	0.045	0.300	0.009	0.783	0.000
	(0.060)	(0.084)	(0.161)	(0.234)				
Hospital stays per year (per 1,000 benes)	1521.801	1505.913	1488.117	-17.796	-1.200	-0.020	0.681	0.000
	(13.895)	(18.885)	(29.884)	(44.809)				
SNF size	204.775	116.487	121.063	4.577	3.800	0.091	0.112	0.004
	(2.505)	(1.292)	(1.684)	(2.967)				
Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock (proportion)	0.151	0.136	0.137	0.001	0.600	0.002	0.906	0.000
	(0.005)	(0.007)	(0.011)	(0.016)				
Diabetes with Chronic Complications (proportion)	0.197	0.188	0.189	0.001	0.300	0.001	0.994	0.000
	(0.005)	(0.007)	(0.013)	(0.019)				
Coagulation Defects/Other Specified Hematological Disorders (proportion)	0.148	0.140	0.141	0.001	1.000	0.004	0.841	0.000
	(0.005)	(0.006)	(0.012)	(0.016)				
Major Depressive, Bipolar, and Paranoid Disorders (proportion)	0.173	0.162	0.169	0.007	4.000	0.018	0.702	0.000
	(0.005)	(0.007)	(0.012)	(0.017)				

Table A.1 *(continued)*

Measure	Potential comparisons (n = 5574)	Matched comparisons (n = 2957)	Treated (N = 901)	Adjusted difference (matched)	Percentage difference (matched)	Standardized difference (matched)	*p*-value (matched)	Equivalence *p*-value (matched)
Cardio-Respiratory Failure and Shock (proportion)	0.195	0.191	0.194	0.003	1.600	0.008	0.859	0.000
	(0.005)	(0.007)	(0.013)	(0.018)				
Congestive Heart Failure (proportion)	0.330	0.319	0.319	0.000	0.000	0.000	0.910	0.000
	(0.006)	(0.009)	(0.016)	(0.022)				
Specified Heart Arrhythmias (proportion)	0.365	0.363	0.370	0.007	2.000	0.015	0.595	0.000
	(0.006)	(0.009)	(0.016)	(0.023)				
Vascular Disease (proportion)	0.245	0.243	0.242	0.000	-0.100	-0.001	0.949	0.000
	(0.006)	(0.008)	(0.014)	(0.020)				
Chronic Obstructive Pulmonary Disease (proportion)	0.214	0.193	0.194	0.001	0.700	0.004	0.942	0.000
	(0.005)	(0.007)	(0.013)	(0.019)				
Acute Renal Failure (proportion)	0.256	0.240	0.246	0.005	2.200	0.012	0.713	0.000
	(0.006)	(0.008)	(0.014)	(0.019)				
Propensity Score	0.129	0.198	0.199	0.001	0.600	0.011	0.781	0.000
	(0.001)	(0.001)	(0.004)	(0.005)				

	Chi-squared statistic	Degrees of freedom	P-value
Omnibus test for balance on matching variables	25.944	34.000	0.838

Notes: Standard errors in parentheses. Standardized difference calculated as the difference in means divided by the treatment group standard deviation. P-values come from a weighted two-sample t-test. The equivalence test comprises two one-sided tests of whether the difference between the treatment and comparison group means exceeds 0.25 standard deviations (in either direction) of that variable. If we can reject that hypothesis, we are confident that the true difference between the two populations being compared is less than 0.25 standard deviations.

OREC = Original Reason for Entitlement Code; ESRD = end stage renal disease; HCC = hierarchical condition category; RUG = Resource Utilization Groups; ADL = Activities of Daily Living Score; CMI = Case Mix Index.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Community Care of North Carolina

April 18, 2018

Molly Lynch (RTI International)
Elliott Liebling (RTI International)
Eva Chang (RTI International)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Enhanced medication management can be an effective tool in improving clinical outcomes and in minimizing health care expenditures for individuals with one or more chronic medical

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Community Care of North Carolina received a 12-month no-cost extension during which time the awardee will continue making incentive payments to pharmacies for attributed patients until the HCIA R2 funding ends.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

conditions. Community Care of North Carolina used funding from HCIA R2 to implement the North Carolina Community Pharmacy Enhanced Services Network (CPESN), which was launched on March 1, 2015. This community-based care delivery and payment model is transforming community pharmacies by motivating pharmacists to focus on improving medication management for the most at-risk patients in addition to their traditional role of dispensing medications. A key component of the CPESN is the integration of pharmacists within a larger care team, including primary care providers, specialty providers, and the extended care team of the patient-centered medical home (PCMH). Toward the end of the third year of the cooperative agreement, the program transitioned from per beneficiary per month (PBPM) and encounter-based payments to value-based reimbursement that ties payment to performance on metrics related to clinical outcomes. Through this arrangement, Community Care of North Carolina is actively testing and paying pharmacies through an alternative pay-for-performance payment model that combines risk and value-based payments for participating pharmacies that have submitted eCare Plans for all high-risk patients enrolled with a pharmacy. To be eligible for the program, individuals must be enrolled in Medicare, in the state's Medicaid fee-for-service (FFS) or S-CHIP program, or they must be dually eligible for Medicaid and Medicare.

The awardee expects the CPESN to reduce hospital readmissions and visits to the ED because pharmacists will be proactively giving patients the medication assistance they need and improving care coordination with providers. Outcome measures include better adherence to medication regimens; better health outcomes, such as fewer exacerbations in patients who have asthma and improved blood sugar management in patients who have diabetes; and a higher degree of patient satisfaction with care. Table I.1 provides additional details on the program.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The CPESN integrates medication management strategies into the interactions between patients and community pharmacists while giving pharmacists the incentive to provide enhanced services, collaborate with a patient's care team, and address gaps in care.
Major innovation	The CPESN changes the way community pharmacy is practiced by incentivizing pharmacists through a value-based approach in which pharmacists move from simply filling and dispensing medication to providing enhanced services that address gaps in patient care by improving medication management for the most at-risk patients.
Program components	• A payment model that provides pharmacies with incentives to provide care to high-risk patients and that addresses gaps in care • An integrated care management and medication management information platform to enhance care delivery by involving pharmacists more directly in patient care • Best practices for delivering the enhanced services in community pharmacy settings and training pharmacists throughout NC so that implementation is statewide • Continuous quality care improvement in all participating pharmacies to establish a more standard, lower-cost approach to care
Target population	Medicaid enrollees or Medicare beneficiaries who have one or more chronic medical conditions treated through medication or who are identified by either a referring physician or the pharmacy itself as needing intervention.
Theory of change/ theory of action	By enhancing the role of pharmacists in the ongoing management of medication for patients with chronic diseases, medical costs can be reduced, and the quality and coordination of care can be improved.

Table I.1 *(continued)*

Program characteristic	Description
Payment model	New risk- and value-based payments combined with PBPM payments for care management/coordination services
Award amount	$15,634,150
Effective launch date	3/1/2015
Program setting	Participating community pharmacies
Market area	Patients served by participating community pharmacies
Market location	NC
Target outcomes	Improved clinical outcomes and a reduction in total annual health care expenditures by at least $30 million by 2017

PBPM = per beneficiary per month.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we conclude that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, Community Care of North Carolina engaged 275 participating pharmacies in the CPESN by the end of Year 3, resulting in 328,806 attributed participants, or 92 percent of its enrollment target, by the end of the cooperative agreement. Second, the awardee reached its hiring goals for Year 3, and there was no turnover in leadership through all three years. Third, the awardee provided extensive training and technical assistance for participating pharmacies, which helped them to implement a complex program that altered the typical pharmacy workflow. Fourth, across the three years of the program, participating pharmacies provided about 28,000 unique comprehensive initial pharmacy assessments (CIPAs) to high-risk patients. Finally, both awardee and pharmacy staff reported a noteworthy, positive change in their patients and an overall positive impact on the delivery of care. These developments include improvements in the efficiency and quality of care and in the services provided to participants, in the ability of pharmacy staff to respond in a timely way to participants' needs, and in their responsiveness to participants' preferences and values.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Community Care of North Carolina's CPESN program, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. Community Care of North Carolina proposed an encounter-based payment model in which participating pharmacies are paid a fee for each care plan submitted for a high-risk patient. The awardee is working with the North Carolina Division of Medical Assistance to incorporate the program into a Medicaid reform effort that is slated to take place in the summer of 2019.

Sustainability plans. Community Care of North Carolina is committed to sustaining its role running the multi-state pharmacy collaborative. The awardee is also committed to facilitating the replication of the program by providing expertise and technical assistance to new pharmacies that want to implement the CPESN. The awardee reported that the state Medicaid program has

agreed to fund the program after the cooperative agreement ends, but the funding will not be realized until 2019. Until then, the awardee will continue to make payments to participating pharmacies by using funding from the cooperative agreement during the no-cost extension period. The awardee is confident that high levels of interest from pharmacies and payers will eventually help to sustain the program, but funding will continue to challenge sustainability until the awardee is able to finalize agreements with payers.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a non-clinician staff survey on their perceptions of the effect of the program on the delivery of care. Examples of non-clinician staff include pharmacists, pharmacy technicians, and administrative staff affiliated with Community Care of North Carolina. The non-clinician staff survey was fielded from July to October of 2016 and achieved a response rate of 100 percent (n = 6). Half of the respondents were primarily in an administrative role, and half were primarily in another non-clinical role. We did not weight the survey samples to adjust for nonrespondents. Because of the small number of respondents, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. For each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we based our assessment, and (3) describe the factors that helped and hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

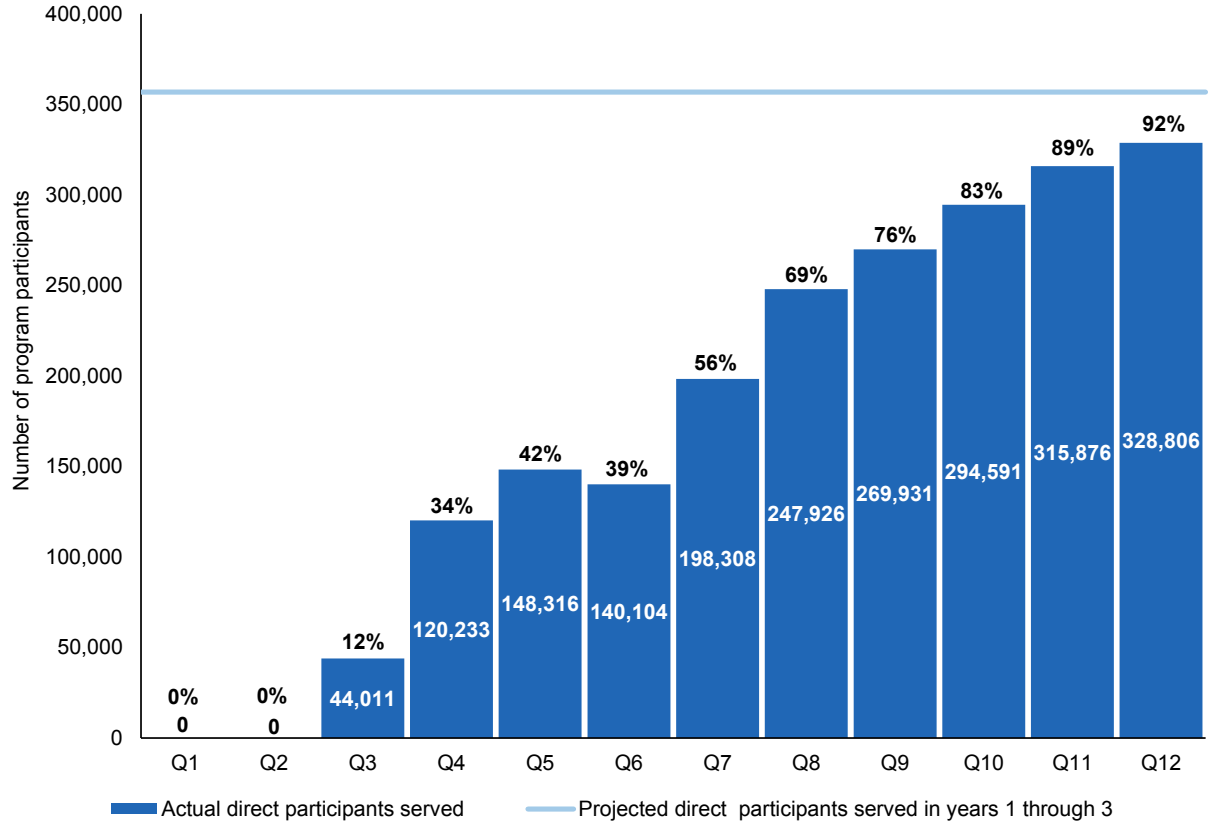
Eligible patients were automatically enrolled in the CPESN through attribution conducted by Community Care of North Carolina. Patients patronizing CPESN pharmacies are what the awardee refers to as the "attribution panel." Using claims data from Medicaid, Medicare, and S-CHIP, Community Care of North Carolina identified these patients from its health information exchange. They included individuals who obtained a prescription from one of the participating pharmacies, have at least one chronic condition, and received 80 percent of their medications from a CPESN pharmacy in the previous 90 days. Community Care of North Carolina developed the algorithm for identifying patients and has revised it over time; however, the criteria for enrollment have stayed the same over the cooperative agreement. For each participant in the panel, Community Care of North Carolina worked with Pioneer Rx, a pharmacy management system vendor, to provide a risk score (for example, risk of hospitalization, complications) that

pharmacists used to identify and prioritize the patients to contact. Assisted by health information technology (IT) systems, participating pharmacies assessed the attributed patients on the basis of their risk status to determine whether they require tailored, enhanced services. Beneficiaries not identified in claims data may enroll if, in the professional judgment of the pharmacist, they need additional medication management services, and they meet the eligibility requirements.

b. Evidence of enrollment effectiveness

Overall, Community Care of North Carolina reported that it enrolled 328,806 direct participants from March 2015 (when it launched its program) through August 2017, which represents about 92 percent of its final three-year projections (Figure II.1). These numbers, which the awardee reports to the implementation and monitoring contractor, reflect the number of Medicare beneficiaries and Medicaid and S-CHIP enrollees attributed to participating pharmacies. In Years 1 and 2, Community Care of North Carolina engaged more pharmacies than anticipated, resulting in higher enrollment of attributed patients. At the conclusion of Year 1, 120,233 direct participants were enrolled; this increased to 247,926 direct participants by the end of Year 2. About 75 percent of all attributed patients were enrolled for at least one year. Community Care of North Carolina raised the number of projected participants to 356,853 from 110,732, a 222.3 percent increase, because more patients were attributed to pharmacies than originally estimated. This number does not reflect the number of participants receiving enhanced pharmacy services, which is described below in Section II.B.2.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017)

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding.

The awardee increased its projected number of participants from 27,954 in Year 1 to 110,732 at the beginning of Year 2 to 356,853 at the beginning of Year 3. The total number of direct participants served and the percentage of the three-year target met are inflated. They reflect the number of attributed participants, not the total number of unique participants who have received services directly from a participating pharmacy. We do not have an estimate of the number of participants who received enhanced services.

c. Barriers and facilitators associated with enrollment effectiveness

Community Care of North Carolina's progress in meeting its three-year enrollment goals was influenced primarily by the rapid growth of the network. The awardee made this growth possible by making the network attractive to community pharmacies—so attractive that it grew faster than anticipated, and in turn, so did the enrollment of attributed patients. Although the awardee exceeded its first enrollment target of attributed participants and almost reached its second enrollment target, a more accurate depiction of enrollment may be represented by the number of CIPAs provided, as they were an indicator of a patient's first encounter with enhanced services. (Section 2 provides additional data on CIPAs.)

2. Delivery of program services

a. Description of and changes to service delivery model

In order to reduce medical costs and improve the quality and coordination of care, Community Care of North Carolina sought to expand the role of pharmacists in the ongoing management of patients with chronic diseases. To that end, pharmacists in the CPESN provided a broad spectrum of enhanced services, but the most intensive service, Community Pharmacy Care Management (CPCM), was essential in achieving the awardee's desired outcomes. The CPCM includes a CIPA, which is both a clinical review of the patient's medical conditions and drug therapies, and a review of the barriers to optimal care, such as low health literacy, social determinants, or a lack of caregiver and/or in-home supports. Critical to CPCM is that the patient receives regular follow-up at the intervals deemed necessary by the pharmacist, who updates the care plan and closely coordinates follow-up activities with other care team members.

As described in previous annual reports, the level of services each pharmacy provided before the CPESN varied a great deal. Some were already offering enhanced services such as home delivery of medications, counseling, linkages with local providers' electronic medical records, and special medication packaging. For other pharmacies, the CPESN represented an extreme change in how they delivered services. Still other pharmacies also offered other enhanced services, including medication reconciliation, home visits, medication delivery, adherence packaging, referrals for behavioral/mental health services and for home and community-based services, and support for improving a patient's self-management and informed decision making. While the type and number of enhanced service offered varied by pharmacy, Community Care of North Carolina set a requirement for participation in Year 3 that a pharmacy must provide at least three CIPAs to remain involved in the CPESN.

One of the major changes in service delivery that came to fruition in Year 3 was a change in the health IT system that was promoted by Community Care of North Carolina. Starting in the early summer of 2017, participating pharmacies transitioned from PHARMACeHOME, a basic pharmacy information exchange platform, to eCare Plans, a shared, real-time plan that documents the process a pharmacist uses to assess and provide care to a patient. eCare Plans, which combines the awardee's final iteration of the payment model with an improved health IT system, can be incorporated into a pharmacist's workflow via a range of platforms. The patient's needs determine when the pharmacist updates the plan, and it can be shared with other providers in a patient's care team. Pharmacists use eCare Plans in two ways: (1) to understand a patient's prescription history in order to deliver effective medication management services and (2) to support the coordination of care among care team providers. The awardee launched eCare Plans and the final payment model on June 1, 2017. Twelve pharmacies were using eCare Plans when the payment model began in June; the number increased to 45 pharmacies in July and 65 in August. Additional participating pharmacies will transition to eCare Plans during the no-cost extension period.

b. Evidence of service delivery effectiveness

Overall, Community Care of North Carolina delivered services effectively, especially in terms of motivating pharmacies to complete CIPAs for participants. However, there were substantial differences in service delivery effectiveness across pharmacies, which reflects the

complex nature of the intervention rather than the awardee's efforts to engage and train the pharmacies. For instance, after the pharmacies had participated in the network for three years, it became apparent that some of them excelled at implementing the program while others floundered. Those that excelled did so because they adapted to the necessary changes in pharmacy workflow, communicated well with other health care providers, and received regular referrals. Other pharmacies—especially those with few resources—struggled to change their workflows, which is a critical step in identifying participants and providing them enhanced services. Furthermore, changes to the incentive payments—including lower payments for many pharmacies during Year 2—led some pharmacies to become less engaged and offer fewer services.

Delivery of intervention services. Participating pharmacies delivered enhanced services to participants, but the number of participants who received CIPAs was a fairly low proportion of total attributed participants. According to one indicator, about 28,000 unique participants across all participating pharmacies have received enhanced services through a CIPA. This figure is almost 70 percent of the awardee's target of providing 40,000 participants with CPCM (CIPA with follow-up coordinated with the rest of the care team). In Year 3, CPESN pharmacists completed 9,892 CIPAs for unique participants. In addition, 23.3 to 38.4 percent of the pharmacies completed one or more CPCM reviews in any given month. We heard anecdotally that pharmacies implemented a suite of enhanced services and follow-up with participants in creative and effective ways. By being so connected to the community, these pharmacies were able to sense when high-risk patients had additional needs such as assistance with meals, so the pharmacies provided referrals or services to fill these gaps.

Staffing and training. Overall, Community Care of North Carolina was able to effectively staff its organization to support the program. The awardee achieved its hiring and training goals by Year 3. As of quarter 12, it has hired 7.1 new full-time employees, which is 100 percent of the Year 3 target. These employees were mostly IT specialists and a small number of management or administrative staff to support data analytics, which has been a staffing need in past years. It is noteworthy that the retention rate of program leaders was 100 percent across all three years of the cooperative agreement. As described in the second annual report, the rapid growth of the CPESN strained the awardee's capacity to provide ongoing, hands-on support and management for a network that nearly doubled in size. Community Care of North Carolina was able to mitigate these growing pains by shifting the type of support offered to pharmacies in a way that aligned the workload with the funded staff.

Community Care of North Carolina trained 1,198 pharmacy staff since project inception, which is nearly 400 percent of its three-year training target. The types of training and technical assistance evolved over the three years to meet the needs of the growing network in innovative ways. Training and engagement methods are described below.

Recruitment and engagement of providers. Community Care of North Carolina was effective in recruiting and engaging pharmacies over the course of the cooperative agreement. As mentioned, 275 pharmacies participated in the CPESN, and program staff reported that the network grew faster than anticipated through word-of-mouth. Once a part of the network, Community Care of North Carolina engaged the pharmacies through regional trainings, peer-to-peer mentorship, and a variety of individualized types of technical assistance, including

workflow assessments conducted by the awardee's university partner (NC State). By Year 3, some pharmacies had not completed any CIPAs, which is considered a marker of low engagement. The awardees asked these pharmacies to submit an action plan if they wanted to continue participating in Year 3, and many of them were able to produce a plan and stay involved in the project. As of August 2017, 65 pharmacies were using eCare Plans. Community Care of North Carolina anticipates that more pharmacies will transition to the no-cost extension, but the network may be smaller overall.

Engagement of program participants. Program participants were engaged at the pharmacy level, particularly those who received CIPAs. Because pharmacists tailored enhanced services to each high-risk patient based on information in the CIPA, the program was able to meet the needs of patients. The majority of non-clinical staff surveyed agreed that they successfully engaged participants in the program. In our interviews, numerous pharmacists talked in anecdotal terms about reaching and engaging high-risk patients by changing their pharmacy workflow and providing enhanced services. For example, a pharmacist read about his patient's spouse's death in the local news and was able to provide follow-up around the loss of that patient's main caregiver. Many pharmacists reported on the impact of medication management and coordinated care on these patients, which is described more fully in Section C.

c. Barriers and facilitators associated with service delivery effectiveness

Community Care of North Carolina faced a "Catch-22" that had an impact on service delivery effectiveness. On one hand, the awardee was constantly collecting feedback from participating pharmacies and sought to enhance its intervention based on this feedback and on the lessons it learned from implementing the program. On the other hand, pharmacies reported that every change in incentives and in the payment model disrupted their workflows.

This situation also manifested itself in the health IT platform. The awardee designed PHARMACeHOME to support its care management services and to allow pharmacists to both document the activities involved in providing enhanced clinical services and submit this information to Community Care of North Carolina for monthly payment. However, service delivery was often stymied by PHARMACeHOME, which prevents pharmacists from entering data because of what they continue to consider a complex and time-intensive system. In the third program year, however, Community Care of North Carolina found a solution in eCare Plans, a health IT platform that is integrated into the pharmacists' workflows. It can be incorporated by a range of software vendors and is considered easier to use than PHARMACeHOME. If pharmacists and their staff are given enough training and technological support, eCare Plans should make it easier for them to document services and coordinate care because it can be updated and reviewed by multiple providers.

The facilitators of effective service delivery included the use of incentives for pharmacies through the payment model, the buy-in and enthusiasm of participating pharmacies for the CPESN, the extensive training and technical assistance that Community Care of North Carolina provided for pharmacies, and the perceived positive impact of the program in the minds of awardee leaders and participating pharmacies.

Community Care of North Carolina also streamlined its technical assistance to pharmacies by developing "change packages" that could be given to a pharmacy at any stage with step-by-

step instructions on how to better implement enhanced services. In particular, incorporating CIPAs into the pharmacy workflow was a major challenge. The awardee also provided intensive technical assistance to pharmacies that have struggled the most. The assistance addressed barriers to completing the CIPAs such as entering data into PHARMACeHOME and identifying which patients to target for what services.

Barriers associated with service delivery effectiveness included not having the appropriate pharmacy-level staff, the varying levels of engagement across pharmacies, and the time commitment required to complete CIPAs. For instance, despite adequate training, some pharmacies found it difficult to hire and train staff. The majority of non-clinical pharmacy staff surveyed noted the following barriers: staff turnover/unfilled positions, insufficient staff time for the amount of work required, additional work for staff, issues related to using program technology, lack of a clearly defined workflow, and difficulty working across collaborating organizations.

Service delivery effectiveness was affected by a pharmacy's level of engagement in the CPESN. Program leaders and frontline staff reported that highly engaged, high-performing pharmacies were able to alter their workflows to identify, assess, and provide enhanced services to high-risk patients who needed them. However, other pharmacies in the network struggled with workflow changes, and they have completed few, if any, CIPAs.

Other barriers to effective service delivery were related to the influx of high-risk patients attributed or referred to high-performing pharmacies. Even when these pharmacies effectively changed their workflow to accommodate the intervention, the additional patients strained pharmacy resources and affected payment based on performance ratings, as these high-risk patients often had multiple ED visits while they were being stabilized. Our interviews showed that workflow and service delivery could be improved by adding staff, redefining the staff's roles, and leaders within pharmacies creating a team approach. The pharmacies with a culture of quality improvement and suitable action plans were seen by program leaders as most compatible with the CPESN.

Program staff identified barriers to enrolling participants via encounters such as CIPAs, including the time required for both the pharmacist and the patient, additional requirements such as meetings and additional work related to patient documentation, a lack of shared understanding of the program, and both patient and provider satisfaction with the old system of care. In our interviews, staff also noted that some patients who are high-need and/or high-risk may not have been enrolled in the intervention because of attribution requirements or because they refuse enhanced services. Program staff also indicated that the unexpected growth of the network in Year 2 was somewhat of a stretch for awardee staff in terms of providing hands-on technical assistance to pharmacies, which may have affected how well pharmacies were able to identify high-risk patients. However, by Year 3, Community Care of North Carolina leaders had stabilized the network by providing key benchmarks that participating pharmacies had to meet (for example, providing at least three CIPAs over the past year).

C. Assessment of perceived program effects on the delivery of care and outcomes

During the interviews, frontline staff enthusiastically reported anecdotal "success stories" of participants they had reached through the program. This response aligned with responses in the awardee survey about the CPESN being somewhat or very effective in achieving its goals. The majority of non-clinical staff surveyed also reported that they perceived the CPESN as having a positive effect on care coordination and on the achievement of participants' health goals, satisfaction, and quality of life. All of the non-clinical staff surveyed thought that implementing the program was worth the effort, and nearly all of them would recommend the program to a close friend or relative in need of care.

> "I absolutely think that [the program goals are] attainable. But they're also the cornerstone of the whole motivation of the movement as well. I do expect and I'm hopeful that those types of metrics will be met. And I believe that without seeing any data point whatsoever, I've seen the impact. I'm a community pharmacist. I've seen the impact that we're making in our own pharmacy, as well as meeting with others and learning about their stories and what they do selflessly to take care of their patients. I can't imagine a community without some of these higher-performing pharmacies. So, I would expect there to be successful metrics made or some type of movement of the needle. And those are certainly—the metrics or the performing metrics that you just mentioned are the cornerstone of that."
>
> —Participating pharmacist

One of the primary factors that has driven this positive perception of program effects is the buy-in from community pharmacies. They have seen a clear advantage to participating in the CPESN. They feel this way, in part, because the network provided a model that will allow them to survive in the current environment, but more important, they are enthusiastic about new ways of working that demonstrate a noteworthy benefit for their patients. Multiple non-clinical staff also pointed out the positive impact of the CPESN on the following: the quality and efficiency of care and of services provided to participants; the ability of staff to respond in a timely way to participants' needs, preferences, and values; and access to care and services for all participants.

According to program leaders and frontline staff, the following additional factors that influenced their perception of the program's effect:

- A patient's direct access to the pharmacist as a health care provider and frequent follow-up by that provider is expected to lead to a reduction in ED visits. One pharmacist noted that high-risk patients may see their primary care provider every three months, but they can access their pharmacist every week.
- Frontline staff repeatedly reported that they clearly saw the program's positive impacts on their high-risk patients. Most staff noted that they observed better adherence to medications. Others reported receiving letters from specialists and other health care providers describing the benefit they had seen in patients who have gained control of their disease. Providers in the community have noticed a difference between the high-risk patients who use a CPESN pharmacy and those who use a traditional retail pharmacy.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Community Care of North Carolina effectively implemented a program that has the potential to result in the desired outcomes as described in the theory of action. The awardee successfully engaged pharmacies in the CPESN and attributed patients to these pharmacies. It also provided training and technical assistance to pharmacies in a variety of ways to help them change their workflows and provide enhanced services to the most high-risk patients. Anecdotally, frontline staff we spoke with told us they have seen impressive changes in their patients as a result of the enhanced services they received through the CPESN. However, some findings from the implementation evaluation may have implications for the impact evaluation. Although program leaders and frontline staff view the CPESN favorably, it may be difficult to see evidence of cost savings or improved health outcomes because of many factors, as explained below.

- Despite the awardee's success in engaging pharmacies and enrolling patients, it is important to note that enrollment reflected the number of Medicare beneficiaries as well as Medicaid and S-CHIP enrollees who were attributed to participating pharmacies, not the number of participants who received enhanced pharmacy services. As such, the group in which we would be most likely to see a change is relatively small. It consisted of patients who were identified as high risk at the pharmacy level and received a CIPA. Those who received a CIPA, continual follow-up, and coordinated care will likely see a reduction in their health care costs. However, there may be a smaller effect across all attributed patients. A subgroup analysis should be conducted to account for participants who received specific types of enhanced services.
- Pharmacies varied widely in how they delivered services. As a result, fidelity and pharmacy engagement can be difficult to measure and categorize. Not all pharmacies identified as "high engagement" pharmacies implemented all components of the program. This may wash out the overall program effect in pharmacies that did not implement major changes. A subgroup analysis may be needed to observe changes in specific patient cohorts (for example, patients taking chronic pain medication).
- Multiple changes to the payment model over time created different incentives that affected how pharmacies delivered services. For example, as Community Care of North Carolina moved toward eCare Plans and away from CIPAs in Year 3, some pharmacies reported that they did not invest as much time in providing CIPAs during this transition period as they were waiting for the final alternative payment model. It may be difficult to control for this change in the impact evaluation.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Community Care of North Carolina's CPESN program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Community Care of North Carolina received a 12-month no-cost extension and will continue to operate until August 31, 2018, but enrollment of new participants will end February 28, 2018. The projected sample sizes in Table III.1 are based upon actual enrollment as of August 31, 2017, projected through February 28, 2018, to allow for all participants to receive six months of program exposure, a requirement for inclusion in our evaluation to minimize dilution of program effects due to limited exposure to the program components. Due to processing lags in Medicaid data, we have not confirmed that all 84,896 Medicaid beneficiaries meet program eligibility for inclusion in our impact evaluation.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: Community Care of North Carolina

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure, based on enrollment through August 31, 2017	72,001[a]
Projected Medicaid population with 6 months of program exposure, based on enrollment through August 31, 2017	84,896[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	3,081
Likelihood of all-cause hospitalizations	2,599
MDE sample size requirement to detect 20% effect	
Total expenditures	770
Likelihood of all-cause hospitalizations	650
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline

Table III.1 *(continued)*

Evaluability domain	Response
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	Some issues, but probably surmountable; expect to select a comparison group
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	Intervention components delivered reflecting greater intensity of intervention

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We anticipate conducting a rigorous impact evaluation of the awardee's program by using difference-in-differences estimation with propensity score matched comparison groups of pharmacies for Medicare and Medicaid beneficiaries who received services from the treatment and comparison group pharmacies. However, due to attribution methodologies for the treatment group that require pharmacy data, an analysis of program effects for Medicare beneficiaries will be limited to the approximately four-fifths of Medicare patients who have Part D benefits. Challenges for the evaluation include the changing treatment status of pharmacies and the possible unobserved differences in capabilities between participating pharmacies and comparison group pharmacies. We have sufficient participation by Medicare and Medicaid beneficiaries to detect small effects (less than 10 percent) on Medicare and Medicaid expenditures. Community Care of North Carolina received a 12-month no-cost extension.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents our summary of the baseline characteristics of Medicare fee-for-service (FFS) beneficiaries who are participating in Community Care of North Carolina's program. We measured the characteristics during the 12 months before each beneficiary's enrollment in the intervention. Participants are enrolled monthly into the Community Care of North Carolina program through two methods: (1) the patient is attributed to the pharmacy (because the patient has one or more chronic medications filled by an intervention pharmacy and 80 percent or more of those medications are filled there in the previous 90 days) or (2) the pharmacist identifies and documents that the patient has a drug therapy problem (DTP). The majority of participants are enrolled passively through attribution to a CPESN pharmacy. For the purpose of our future impact evaluation, the treatment group will consist of individuals with Medicare FFS and Medicare Part D,[4] Medicaid, or both who were enrolled in Community Care of North Carolina's intervention. For this summary, we include all Medicare FFS beneficiaries, including those without Part D, to describe the baseline characteristics of all Medicare FFS

[4] Because prescription data are only available for Medicare FFS beneficiaries with Part D, treatment group beneficiaries for the Medicare population only include Medicare Part D beneficiaries for consistency with the identification of the comparison group.

patients who are potentially eligible for the treatment group. Due to lags in the receipt of Medicaid data, we were unable to present the baseline characteristics for the Medicaid treatment group in this report.

Community Care of North Carolina began enrolling beneficiaries in March 2015. As of April 30, 2016, the awardee had enrolled 144,257 unique participants in the program by using one of the enrollment methods described above.[5] There were 57,158 Medicare participants and 94,163 Medicaid participants (33,686 individuals participated in both Medicare and Medicaid). The remaining 18 percent (26,622 individuals) either had other sources of health care coverage or they were uninsured; these participants were not included in our analysis. Fewer Medicare-only participants are enrolled in Community Care of North Carolina because pharmacy fill data are being used to identify attributed Medicare patients. Fill data are not available for all participating pharmacies.

In presenting the baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they must have been enrolled in the awardee's program on or before May 31, 2016, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. After we excluded patients who did not meet the above criteria, a total of 40,480 Medicare FFS participants were included in the analysis of baseline characteristics for this report.

The Medicare FFS beneficiaries who are participating in the Community Care of North Carolina's program are a diverse group, based on demographic and health status characteristics (Table III.2). More than two-fifths (41 percent) of beneficiaries are younger than 65, while 10 percent are older than 85. Participants are more likely to be female (62 percent). In addition, most participants are predominantly white (62 percent) or black (34 percent), which reflects the racial composition of the Medicare-Medicaid dual eligible and Medicare-only beneficiaries in North Carolina.[6] Compared with 16 percent of Medicare beneficiaries nationwide and 19 percent of Medicare beneficiaries in North Carolina,[7] 55 percent of participants were originally eligible for Medicare because of a disability. One percent of participants were entitled to Medicare because of end-stage renal disease (ESRD). Thus, Community Care of North Carolina is recruiting a population that generally has significant health care needs and high Medicare

[5] For the analysis, participants were considered enrolled after they were first attributed to a participating pharmacy or were identified and documented as having a DTP by a pharmacist in a participating pharmacy.

[6] North Carolina's Medicare population is approximately 74 percent white and 20 percent black, while its dual eligible population is approximately 53 percent white and 36 percent black. See Kaiser Family Foundation analysis of the March 2015 Current Population Survey Annual Social and Economic Supplement and analysis of fiscal year 2011 MSIS and CMS-64 reports.

[7] See Kaiser Family Foundation, "Distribution of Medicare Beneficiaries by Eligibility Category in 2013." Available at http://kff.org/medicare/state-indicator/distribution-of-medicare-beneficiaries-by-eligibility-category-2/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D.

expenditures. More than 70 percent of participants are Medicare-Medicaid dual eligibles, a designation that signifies a high level of social need. Participants have a mean hierarchical condition categories (HCC) risk score of 1.47 (relative to a national mean risk score of 1.00), which indicates poorer health status and greater needs for care than the general Medicare FFS population.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Community Care of North Carolina's program through April 30, 2016

	All participants (N = 40,480)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	16,673	41
65 to 74	12,205	30
75 to 84	7,746	19
85 and older	3,856	10
Gender		
Female	25,217	62
Male	15,263	38
Race		
White	25,108	62
Black	13,910	34
American Indian, Alaska Native, Asian/Pacific Island American, or other	933	2
Hispanic	336	0.83
Original reason for Medicare eligibility		
Old age and survivor's insurance	17,875	44
Disability insurance benefits	22,135	55
ESRD[a]	470	1
Hospice[b]	254	0.63
Medicare/Medicaid dual status, percent dual[b]	29,291	72
HCC score[c]		Statistic
Mean		1.47
25th percentile		0.69
Median		1.09
75th percentile		1.77

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of April 30, 2016

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as either the date on which the program started in a facility (for existing residents) or the date on which a beneficiary first became a resident during the program period. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

Table III.2 *(continued)*

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

DTP = drug therapy problem; ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category.

Consistent with the higher HCC score, participants had higher expenditures in the year prior to enrollment. In Table III.3, we report baseline utilization and expenditure data for a common set of measures, including the four CMMI core measures. The Community Care of North Carolina program expects to reduce medical expenditures by reducing (1) ED visits, (2) complications from poor control of chronic conditions, (3) and crisis management among mentally ill (and medicated) patients. We examined baseline cost of care by calculating average PBPM Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $978 (relative to the North Carolina Medicare FFS beneficiary average of $719 in 2014).[8] The quarterly average PBPM ranged from $925 to $1,052. Average PBPM Medicare payments for acute inpatient services ($322), physician services ($258), and outpatient services ($195) were the largest drivers of the total cost of care. Quarterly expenditures for these services were relatively stable over time, although all average costs were higher in quarter 4 of the baseline year. Payments for acute inpatient services ranged from $302 to $357 PBPM; for physician services, from $249 to $267 PBPM; and for outpatient services, from $184 to $206 PBPM.

[8] Unless otherwise noted, national and state data in this paragraph is from the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2016. Available at http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf?sfvrsn=0.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Community Care of North Carolina's program through April 30, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	40,480	39,037	39,736	40,465	40,480
Average Medicare expenditures PBPM[a]					
Total	978 (10)	925 (14)	939 (15)	990 (15)	1,052 (15)
Acute inpatient	322 (5)	306 (8)	302 (9)	320 (10)	357 (9)
Inpatient other[b]	49 (3)	48 (4)	48 (5)	47 (3)	53 (6)
Outpatient[c]	195 (3)	184 (3)	191 (4)	199 (4)	206 (4)
Physician services	258 (2)	250 (3)	249 (2)	264 (3)	267 (3)
Home health	43 (1)	41 (1)	42 (1)	45 (1)	46 (1)
Skilled nursing facility	60 (2)	55 (3)	60 (3)	64 (3)	62 (3)
Hospice	13 (1)	7 (1)	10 (1)	14 (1)	21 (1)
Durable medical equipment	37 (1)	35 (1)	37 (1)	38 (1)	39 (1)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	381 (5)	364 (8)	360 (7)	380 (7)	418 (8)
Outpatient ED visits	1,036 (12)	1,026 (17)	1,033 (16)	1,041 (16)	1,041 (15)
Observation stays	99 (2)	96 (3)	96 (3)	101 (3)	101 (3)
Primary care visits in any setting	7,791 (39)	7,584 (51)	7,649 (52)	7,811 (51)	8,099 (53)
Primary care visits in ambulatory settings	6,297 (28)	6,160 (36)	6,218 (36)	6,330 (36)	6,465 (36)
Specialist visits in any setting	9,169 (55)	9,079 (74)	9,006 (70)	9,237 (72)	9,337 (71)
Specialist visits in ambulatory settings	7,016 (39)	6,778 (48)	6,888 (48)	7,137 (48)	7,240 (47)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	22 (< 0.5)	7 (< 0.5)	7 (< 0.5)	8 (< 0.5)	8 (< 0.5)
Percentage with an outpatient ED visit[e]	41 (< 0.5)	16 (< 0.5)	17 (< 0.5)	17 (< 0.5)	17 (< 0.5)
Percentage with an observation stay[f]	8 (< 0.5)	2 (< 0.5)	2 (< 0.5)	2 (< 0.5)	2 (< 0.5)
Percentage with a 30-day readmission among all discharges	18 (< 0.5)	16 (1)	18 (1)	19 (1)	18 (1)
Percentage of participants with a readmission among all participants	4 (< 0.5)	1 (< 0.5)	1 (< 0.5)	1 (< 0.5)	1 (< 0.5)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of April 30, 2016

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

The rate of acute care hospitalizations was 381 per 1,000 Medicare FFS participants per year during the baseline year—higher than the North Carolina average in 2014 of 268 per 1,000 Medicare FFS beneficiaries.[9] Twenty-two percent of participants had at least one hospitalization during the year before enrollment. The percentage of discharges with a 30-day readmission among participants (18 percent per discharge) in the baseline year was the same as the North Carolina average in 2014 for Medicare beneficiaries (18 percent per discharge). The rate of outpatient ED visits (1,036 per 1,000 participants in the baseline year or 41 percent of participants) was more than double the 2013 national rate of 454 per 1,000 Medicare FFS beneficiaries.[10] The considerably higher rates of outpatient ED visits suggest that there is an opportunity to reduce use of outpatient ED visits through improved patient self-management and coordination across providers. The rate of ambulatory observation stays was 99 per 1,000 beneficiaries per year in the baseline year. Overall, observation stays at baseline were higher than the national average of 58 per 1,000 beneficiaries per year in 2014.[11] At baseline, the rate of primary care visits in any setting was 7,791 per 1,000 Medicare FFS beneficiaries per year. This rate falls to 6,297 per 1,000 Medicare FFS beneficiaries per year when restricted to ambulatory settings. The rate of specialty care service use in any setting was 9,169 per 1,000 Medicare FFS beneficiaries per year and the rate falls to 7,016 per 1,000 Medicare FFS beneficiaries per year when restricted to ambulatory settings. As with expenditures, the utilization for most services is highest in quarter 4 of the baseline year. Overall, enrolled beneficiaries had higher expenditures, a higher rate of acute care hospitalizations, and a higher rate of outpatient ED visits relative to the national and North Carolina averages for all Medicare FFS beneficiaries—suggesting that there is potential to make improvements in care.

[9] Unless otherwise noted, national and state data in this paragraph is from the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2016. Available at http://www.medpac.gov/documents/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf. Accessed October 2016.

[10] National outpatient ED rate calculated from the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2016. Available at http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf?sfvrsn=0. Accessed August 2016.

[11] See the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2016. Available at http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf?sfvrsn=0. Accessed August 2016.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Toward the end of Year 3, Community Care of North Carolina proposed a third iteration of its payment model, an encounter-based payment model in which participating pharmacies are paid a fee for each care plan submitted for high-risk patients. The awardee is working with the North Carolina Division of Medical Assistance to incorporate the program into a Medicaid reform effort that is slated to take place in the summer of 2019.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Under Community Care of North Carolina's proposed encounter-based payment model, pharmacies receive payment for submitting an initial CIPA and for any subsequent monthly updates. This payment is in addition to the dispensing fee that pharmacies already receive for filling prescriptions. There are no penalties for not submitting an updated care plan each month apart from failing to receive the monthly fee. Pharmacies must meet the participation requirements established by Community Care of North Carolina in order to be eligible for these payments.

The monthly fee ranges from $2.50 to $40, depending on the patient's risk score[12] and on the pharmacy's performance score on risk-adjusted metrics (Table IV.1). The patient's risk score is based on the risk for hospitalization and for needing a detailed consultation with a pharmacist. Community Care of North Carolina, through its role as a Medicaid primary care case management system, has access to Medicaid claims that it can use to calculate risk scores for Medicaid enrollees. However, the awardee does not have real-time access to Medicare claims, so it uses a third-party vendor to estimate risk scores for Medicare beneficiaries. Similarly, Community Care of North Carolina uses Medicaid claims to calculate a performance score for each pharmacy based on attributed Medicaid enrollees, but it cannot do this for Medicare beneficiaries. Pharmacies can earn up to 11 points based on the following risk-adjusted performance metrics, which are updated quarterly:

- Total cost of care (3 possible points)
- ED utilization (2 possible points)
- Inpatient hospital utilization (2 possible points)
- Medication adherence (4 possible points across 4 measures)

[12] Community Care of North Carolina provides a risk score for attributed patients to pharmacy records so that pharmacists can more easily identify high-risk patients who need enhanced services.

Table IV.1. Potential monthly fees for care plans based on pharmacy performance scores and patient risk

Year 3 CCNC alternative payment model				
	Pharmacy's most recent performance score			
Patient's risk score	Above average (8–11 points)	Average (6–7 points)	Below average (4–5 points)	Review for network inclusion (0–3 points)
≥ 85	$40	$30	$20	$10
75–84	$30	$22.50	$15	$7.50
60–74	$20	$15	$10	$5
50–59	$10	$7.50	$5	$2.50
< 50	$5	$3.75	$2.50	$1.25

Source: Provided by Community Care of North Carolina (CCNC).

C. Status of the payment model

The payment model proposed by the awardee is its third iteration; the previous iterations were described in the first and second annual reports. The current model builds on lessons learned by Community Care of North Carolina in attempting to implement the previous iterations. The awardee began implementing this model at the end of the cooperative agreement and used HCIA R2 funds to provide a group of pilot pharmacies with the monthly fee for submitting the CIPA and subsequent updates for Medicaid enrollees or Medicare beneficiaries. The awardee will test this iteration in the final year of the program as allowed through the no-cost extension.

Because Community Care of North Carolina focused on Medicaid enrollees, the payer most likely to adopt its program is the state Medicaid program. The awardee has worked closely with the North Carolina Division of Medical Assistance to incorporate the program into a Medicaid reform effort that is slated to take place in the summer of 2019. Community Care of North Carolina also included the program as an option for a prepaid health plan in a Medicaid waiver submitted to CMS and included in a white paper on North Carolina's vision for Medicaid reform. Until a decision is made, Community Care of North Carolina is working with the Division of Medical Assistance to establish a bridge that would allow the participating pharmacies to continue the program during the period between the end of the cooperative agreement and the implementation of Medicaid reform.

There is no mechanism that Community Care of North Carolina can use to bill Medicare for services provided to beneficiaries. The awardee is in the early stages of exploring whether it is possible for pharmacies to establish a legal and financial relationship with physicians to provide services and share in the reimbursement.

D. Factors associated with the development of the payment model

Community Care of North Carolina acknowledged that its payment model has required pharmacies to make significant changes to the way in which they have typically operated. Specifically, performance measures and value-based purchasing models are increasingly

common throughout the health care system; however, retail pharmacies have continued to rely on a reimbursement model that is based on dispensing fees and drug costs with few, if any, pharmacy-specific performance measures. Pharmacists that were interested in providing care management services to their patients, such as medication reconciliation and referrals to community resources, did not have a vehicle for receiving remuneration for these efforts. By creating a payment model that both rewards pharmacists for the comprehensive services they are providing and adjusts their fees based on performance, the awardee is moving pharmacy reimbursement toward a value-based payment model. In creating this model, the awardee has worked closely with pharmacies to identify the components of the model that are most important, including establishing a price point below which pharmacies ceased to participate.

The awardee also acknowledged that risk adjustment of the performance measures has been critical in developing the payment model, especially because patients with complex medical conditions have sought out the high-performing participating pharmacies in order to take advantage of the care management services that are offered. Because these pharmacies therefore experienced adverse selection by attracting a disproportionate number of high-risk patients, they ended up with lower performance scores. Although the awardee has worked closely with clinical experts and statisticians to identify key risk adjustment factors that are currently in place, it acknowledged that there are probably other factors that it cannot address in its risk adjustment to account for health literacy and social determinants of health, such as transportation and other issues. These are the issues that the high-performing pharmacies are addressing through services, such as home delivery of medications, that lead high-risk patients to seek out their services.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Community Care of North Carolina is committed to sustaining its role running the multi-state pharmacy collaborative. It is also committed to facilitating the replication of the program by providing expertise and technical assistance to pharmacies that are implementing the program. The awardee reported that it has succeeded in engaging the state Medicaid program to fund the program after the cooperative agreement ends, but the funding will not be realized until 2019. Until then, the awardee will continue making payments to participating pharmacies by using funds from the cooperative agreement during the no-cost extension period. The awardee expressed confidence that high levels of interest from pharmacies and payers will eventually help to sustain the program, but funding will continue to be a challenge to sustainability until the awardee finalizes agreements with payers.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Community Care of North Carolina had developed a multistate pharmacy collaborative with interested organizations representing community pharmacies in other states across the country. The collaborative fosters sustainability and

replication of the program in pharmacy networks in a variety of states. By the end of the second program year, 17 states had started implementing the program, and an additional 18 states had expressed an interest in doing so. As described in Chapter IV, Community Care of North Carolina had also implemented the first iteration of the payment model as of the second program year, and it planned to roll out the next iteration in the third program year. Finally, the awardee attempted to provide technical assistance to participating pharmacies more efficiently by giving the pharmacies "change packages," which included instructions on how to implement the program.

C. Implementing the SSR plan: progress and changes

Sustainability. Community Care of North Carolina reported that it plans to use the no-cost extension to continue to support pharmacies in the CPESN, although awardee leaders reported that there is uncertainty about whether pharmacies that participated in the program would continue providing services after the funding ends. The awardee will stop payments to pharmacies after funding from the cooperative agreement ends, which will occur at some point before August 31, 2018, the end date for the no-cost extension. However, awardee leaders reported that they plan to continue paying participating pharmacies through the alternative payment model during the no-cost extension until the funding is depleted. In addition to payments, awardee leaders also emphasized that pharmacies need non-financial support, such as technical assistance provided by the awardee, to sustain the program. "There is real transformation work to be done [in pharmacies in order to sustain the program successfully]," an awardee leader explained.

> "Funding is the biggest barrier. We don't know if pharmacies will want to continue these services without a payment."
>
> —Awardee leader

Scalability. Community Care of North Carolina did not report that it plans to scale its program.

> "We want to continue to disseminate and package support to those local networks.... We're feeling very confident that [we have] a business plan with some longevity."
>
> —Program leader

Replicability. Community Care of North Carolina used HCIA R2 funding to replicate its program through a multistate pharmacy collaborative that included community pharmacy stakeholders across the country. By the end of the third program year, 40 states had requested the awardee's assistance to support the establishment of a CPESN in their areas. Awardee leaders said that they are committed to continuing to provide technical assistance to pharmacies locally and nationwide for the next few years not only to make it easier for pharmacies to sustain the program but also to help new pharmacies replicate the program.

Awardee leaders also reported that they plan to obtain clinically integrated network status, and charge participating pharmacies a fee for its services. Services include technical assistance and subject matter expertise on topics such as quality improvement, creating a medium for fostering partnerships between pharmacies and other care team members, marketing the program's value to stakeholders, and identifying more high-performing pharmacies. One

program leader summarized the plan to continue supporting CPESN pharmacies as doing "whatever the network needs in order to be a successful local network and self-sustaining."

D. Factors associated with progress toward implementing the SSR plan

Community Care of North Carolina reported that pharmacies have shown a high level of enthusiasm for and interest in the CPESN. This response may have resulted from increased interest in value-based payment models in health care in general and by payers in particular. In the words of one interview respondent, "[The pharmacies are] excited about being partners in health reform and offering value-based care. Quite honestly, they are very accepting that payment models may evolve to a value-based [model]." In addition, awardee leaders said that the program is likely to be sustained because payers are highly interested in value-based payments. "Rapidity of payment reform on the medical side is really driving a lot of the behavior change [for pharmacies]," one awardee leader posited. As discussed in Chapter IV, the awardee has already engaged North Carolina's Medicaid program, and awardee leaders believed that the program would continue to easily engage payers, especially other state Medicaid programs. "Medicaid is probably the most promising because the population matches these types of interventions. Also, many states are looking for solutions and writing it into the waivers and state plans."

Despite the potential for engaging payers, Community Care of North Carolina reported that funding is the biggest challenge to sustainability. As mentioned, the awardee is not sure that participating pharmacies will continue to provide program services after payments end. In addition, pharmacies in other networks are not paid, and awardee leaders reported that they are concerned that those pharmacies may burn out. The informal leaders of networks in other states work on CPESN implementation in their community in addition to their full time job, awardee leaders explained. "They will easily get depleted over time," one awardee leader said, adding that one solution to burn out is to provide these network leaders with more financial resources.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension that allows it to continue to enroll new beneficiaries. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Community Care of North Carolina's CPESN program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicare beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Catholic Health Initiatives Iowa Corp., DBA Mercy Medical Center–Des Moines

April 18, 2018

Allison Steiner (Mathematica Policy Research)
Lee-Lee Ellis (Mathematica Policy Research)
Boyd Gilman (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Catholic Health Initiatives was awarded a six-month no-cost extension, and will continue to enroll participants in the program through February 28, 2018.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

The Mercy Accountable Care Organization (ACO), which is headquartered in Des Moines, Iowa, and is a division of Catholic Health Initiatives Iowa Corp. (CHIIC), partnered with 25 rural hospitals affiliated with the Mercy Health Network to implement the Transitioning a Rural Health Network to Value-Based Care program. This program expanded innovative population health activities—such as health coaching at primary care clinics, a disease registry throughout the network, and quality improvement projects at hospitals and clinics—to rural, low-income communities across the state that had higher rates of diabetes, obesity, and disability compared to the rest of Iowa or the United States. Awardee leaders developed this program because they recognized that significant gains could be made in improving population health in poor, rural Iowan communities located in the Mercy Health Networks' three geographically defined chapters: Central Iowa (12 hospitals), North Iowa (9 hospitals), and Siouxland (located in western Iowa and eastern Nebraska, with 4 hospitals).[4] Each network hospital has one or more affiliated primary care clinics where the program's health coaches deliver program services.[5] An innovative feature of the program was that it incorporated the rural critical access hospitals (CAHs) into Mercy ACO's existing Medicare shared savings program (MSSP).

The program had three components (Table I.1). The first was a health coaching intervention implemented at the participating primary care clinics. The health coaching intervention provided short-term, intensive care management for adults with one or more chronic diseases including diabetes, hypertension, chronic obstructive pulmonary disease (COPD), and cardiovascular disease. The health coaching component was expected to improve health outcomes, reduce unnecessary use of services, and lower the average cost of care. At each of the group of clinics affiliated with a given hospital, one nurse health coach and one provider champion implemented the health coaching program.[6] One rural market manager in each of the three regions oversaw the health coaches and the provider champions, and also coordinated communication between the Mercy ACO and the staff at the implementing sites.

A second practice-level component was focused on continuous quality improvement and implemented within each group of hospitals and their affiliated clinics. It used Lean Process Improvement principles to help reduce costs and optimize operations.[7]

Finally, the program included a health information technology (health IT) component. The Mercy ACO used its HCIA R2 funding to create a Datashop platform. The platform was designed to connect data feeds between participating sites and the Mercy ACO, host a disease registry that tracks billing and clinical information for the ACO's patient population, and provide data for quality improvement activities. We describe these characteristics in detail in Chapter II.

[4] The hospitals included 23 critical access hospitals and two medical centers with a rural primary care network.

[5] The main clinic is usually co-located on the hospital campus and any others are in outlying communities.

[6] The health coaching intervention is the primary program component (these are the direct participants as defined by the awardee); however, the health coach also performs additional population health services such as annual wellness visits and transitional care management calls at most sites (Section II.B)

[7] The quality improvement component of the program is a secondary component, and it will not be the focus of the evaluation.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The program uses (1) health coaches to help participants manage their chronic diseases and set and achieve health goals, coordinate patient- and family-centered care services, and connect participants to community-based resources; (2) health IT to facilitate patient identification, care management, and outcome reporting; and (3) quality improvements to reduce costs and optimize operations.
Major innovation	CHIIC expanded existing urban-based population health activities to Medicare and Medicaid beneficiaries in resource-limited, rural communities. In addition, the program was an early adopter of bringing rural practices into an MSSP.
Program components	The program had three components: 1. **Health coaching.** Nurse health coaches helped participants set and achieve health goals, provided self-management training, assessed gaps in care, coordinated care, and connected patients to community resources. Health coaches used care management software to track participants' health and appointments. 2. **Quality improvement.** Lean Process Improvement tools were used to identify areas where operations could be improved, processes could be standardized, and costs could be lowered. 3. **Health IT.** The Datashop platform, which includes care management software and a disease registry, was used for patient identification, care management, and outcome reporting.
Target population	Rural residents, primarily those with one or more chronic conditions (diabetes, hypertension, COPD, and cardiovascular disease) and those with a history of ED utilization.
Theory of change/theory of action	Health coaches and staff at rural primary care clinics identified patients in rural communities who would benefit from the health coaching intervention. Once individuals agreed to participate, the health coach performed activities including goal-setting, medication review, depression screenings, cancer screening referrals, and self-management support. The health coaching intervention made patients more accountable for their health and closed gaps in preventive care, which increased primary care utilization, raised vaccination and screening rates, and reduced ED utilization. These outputs were, in turn, expected lead to the outcomes of better health, appropriate health care utilization, and lower costs.
Payment model	Shared savings with ACO
Award amount	$10,170,496
Effective launch date	09/01/2014
Program setting	Primary care clinics affiliated with rural hospitals
Market area	Rural
Market location	Iowa and Nebraska
Target outcomes	• Health outcomes: improved population health as measured by CMS's 33 ACO quality measures • Health care utilization: increased use of primary care, less use of the ED for non-emergency conditions, and fewer preventable hospitalizations • Costs: reduction in the total cost of care for participants

ACO = accountable care organization; COPD = chronic obstructive pulmonary disease; ED = emergency department; IT = information technology; MSSP = Medicare Shared Savings Program.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was partly successful in implementing the health coaching program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, the awardee enrolled 13,044 participants in the health coaching component—65 percent of its final enrollment projection, which was more than 10 times the initial target (1,295)—by the end of the initial cooperative agreement. However, frontline staff applied the eligibility criteria broadly, enrolling not only patients with more than one chronic condition, but also patients with other health goals (such as weight loss and smoking cessation) or risk factors for chronic conditions.

Second, the awardee implemented health coaching services in all regions, but there was variation across regions – and between sites within a region – in how the intervention was implemented, and there were delays getting the program's IT platforms up and running. Third, the awardee achieved its staffing target, reaching its three-year staffing projection by the third program quarter. Fourth, despite variation between implementing regions and sites and initial reluctance by clinicians to refer their patients, the awardee successfully engaged most implementing providers. Finally, most frontline staff said they believed the program successfully engaged participants and had a positive effect on the delivery of care.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Catholic Health Initiatives, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. The awardee's payment model is based on the creation of a rural MSSP ACO for Mercy Health Network CAHs and their affiliated clinics. The awardee would supplement the new shared savings arrangement with CMS with fee-for-service (FFS) billing for specific health coaching services, including transitional care management, advance care planning, and annual wellness visits. By the end of the three-year cooperative agreement, all of the rural hospitals participating in the health coaching program joined the rural MSSP ACO.

Sustainability plans. The Mercy ACO continued making progress on its sustainability efforts, most notably encouraging sites to participate in the rural MSSP ACO and engaging hospital and clinic leaders. CMS also granted the awardee a six-month no-cost extension. The extension will allow the awardee to continue developing the disease registry, which the Mercy ACO uses to track clinical data for its attributed population. The registry is used by the health coaches and clinic staff to identify eligible health coaching patients and performance improvement projects.

The awardee did not plan to scale the program. However, the health coaching program was implemented at all sites that joined the Mercy ACO during the cooperative agreement, and it will be implemented at any sites that join going forward. Finally, although the awardee did not report any plans to replicate the program, it received inquiries from other rural health systems interested in implementing similar programs.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of four data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The survey of non-clinician staff was fielded from July to October of 2016 with a sample of 56 potential respondents (including health coaches) and achieved a response rate of 96 percent. The survey of clinicians was fielded from March to June of 2017 with a sample of 230 potential respondents and achieved a response rate of 64 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

Finally, using the awardee's encounter-level database, we assessed the program's health coaching services to determine the number of participants served at each site and how often each participant used the health coaching services. Health coaches were expected to manually create a record in an electronic care management tracking system for each encounter with a patient.[8] The tracking system's database includes the name, date of birth, and gender of the enrollee, as well as the date and location of the encounter and the mode through which the service was provided (in person or by telephone). We converted the 64,337 encounter-level records into a unique enrollee-level database by matching records on first and last names and date of birth. We limited our analysis to participants with at least one in-person visit with a health coach to align with the awardee's definition of a health coaching participant.[9] To allow for sufficient follow-up time after enrollment, we excluded participants who enrolled after May 2017 (three months before the end of the initial cooperative agreement). Clinics in the Central region transitioned to a new care management tracking system in November 2016, and were unable to provide encounter data to the awardee during the final nine months of the program. Our quantitative assessment of health coaching services is based on 8,595 enrollees.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close

[8] Initially, all regions used the TAV Health care management software to track health coaching encounters and report encounters to the Mercy ACO. During the ninth program quarter, the clinics in the Central region (affiliated with 12 hospitals) transitioned to the Datashop software. Encounters from the Central region after the ninth quarter were not provided in the encounter-level file.

[9] The encounter-level database provided by the awardee included all encounters for population health services provided by health coaches, not just the health coaching intervention.

the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partly successful in implementing the health coaching program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

The health coaching program targeted patients with one or more chronic conditions (diabetes, hypertension, cardiovascular disease, and COPD), as well as patients with a history of ED utilization or hospitalizations. Over the duration of the award, frontline staff applied the eligibility criteria broadly, enrolling not only patients with one or more chronic conditions, but

also patients with other health goals (weight loss, smoking cessation) or those who had risk factors for chronic conditions. In the Siouxland region, clinics focused particularly on offering transitional care management to and enrolling individuals who had recently been discharged from a hospital. Awardee leaders and rural market managers acknowledged that, although all clinics focused on hypertension and diabetes, some sites focused on other areas based on referring providers' interests, community needs, and billing opportunities.

Health coaches and their assistants identified eligible participants by conducting pre-visit chart reviews, generating lists of patients with chronic conditions from electronic medical record (EMR) registries, and reviewing hospitals' administrative lists of patients with ED visits or hospitalizations.

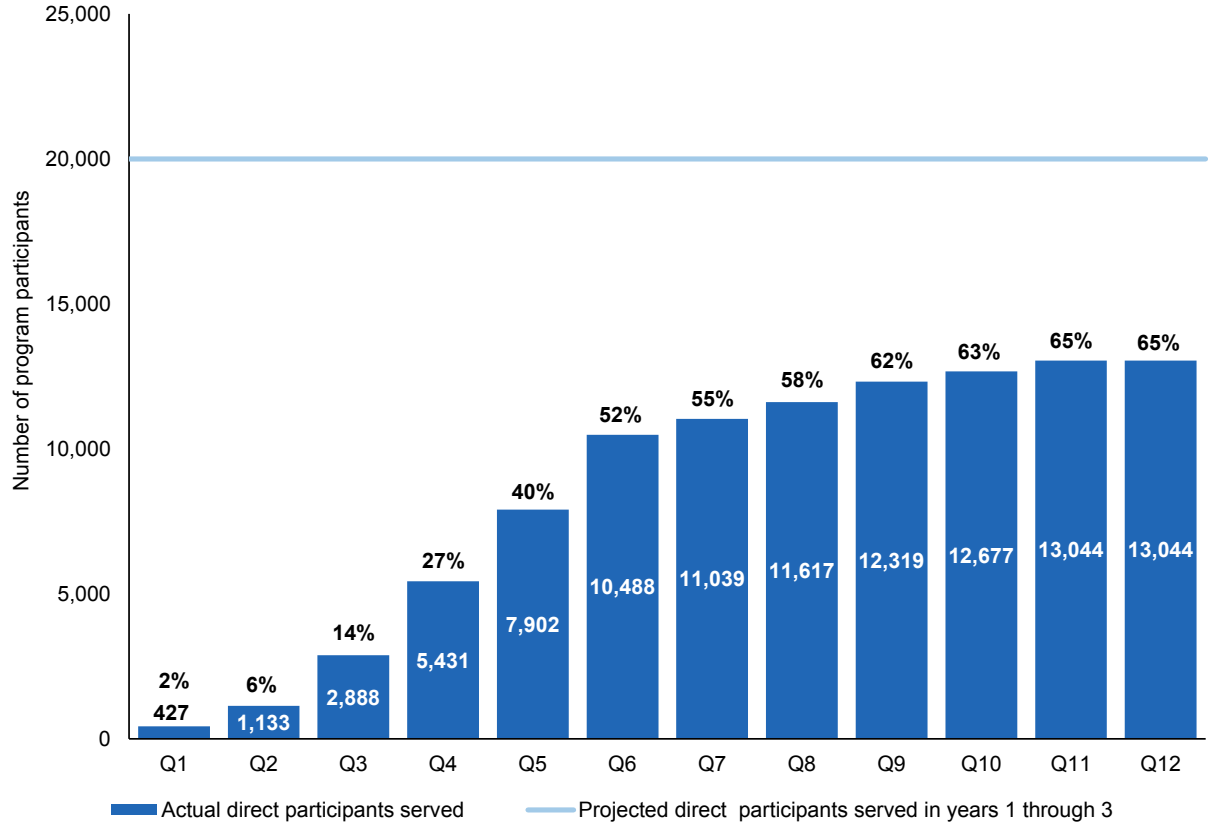
Sites varied in their use of EMR, disease registry, and hospital administrative data to support patient identification. Over the course of the cooperative agreement, some implementing sites transitioned from paper chart reviews to automated identification of eligible patients; this transition did not alter the eligibility criteria, but did make it easier to identify and enroll patients. Patients were also referred by primary care providers (PCPs) and other clinic staff, who identified patients whom they believed would benefit from health coaching services. Once identified as eligible, an individual met with the health coach, who introduced the program. This meeting usually took place when the prospective participant was already in the clinic. If he or she was willing and ready to participate, the health coach and participant used the initial meeting to discuss the individual's goals and decide on an appropriate schedule for their interactions.

b. Evidence of enrollment effectiveness

Overall, the awardee reported that it enrolled 13,044 direct participants into the health coaching intervention from September 2015 (when it launched its program) through August 2017, which represents about 65 percent of its final three-year projections (Figure II.1).[10] The awardee raised its direct participant enrollment projection several times during the three-year cooperative agreement because its initial target was deemed too low. Additionally, the awardee enrolled more participants than originally projected because health coaches were spending less time with each participant than expected. The awardee increased its projection from 1,295 at the beginning of the first program year to 10,000 at the beginning of second, and doubled this to 20,000 at the beginning of Year 3. Therefore, even though the awardee only reached 65 percent of its final three-year projection, we determined that the awardee made significant enrollment progress. In addition, the awardee reported that the number of direct program participants was likely underreported because two care management systems were used to track enrollment, raising reporting issues for some health coaches and causing some difficulty in integrating the two systems. When measured against the three-year projection set at the end of Year 1, the awardee met 130 percent of its projection.

[10] Direct participants are those who actively enrolled in the health coaching program and received health coaching services at one of the participating sites. Direct participant counts do not include individuals who only received other population health services (also provided by health coaches), such as annual wellness visits and transitional care management calls.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee raised its direct participant enrollment projection several times during the three-year cooperative agreement, from 1,295 at the beginning of Year 1 to 10,000 at the beginning of Year 2 to 20,000 at the beginning of Year 3.

c. Barriers and facilitators associated with enrollment effectiveness

The awardee's progress in meeting its three-year enrollment goals was influenced by several key facilitators: (1) integrating health coaches into the care teams, which increased referrals from other clinic staff; (2) identifying health coaching participants through other population health activities; and (3) raising the visibility and acceptance of the health coaching program in the rural communities served by the awardee.

First, rural market managers, who oversee health coaches in each of the three regions, noted that over time, health coaches became better integrated into the care team as other staff members came to fully understand the health coach's role and recognize the value they added. After working closely with the health coaches and seeing their value, other staff were more likely to play an active role in identifying and referring participants to the program. For example, nurses identified eligible patients while refilling prescriptions and reminded PCPs to refer eligible

patients. Similarly, as providers recognized the value of the health coaching services, they were more likely to refer patients and encourage them to participate. When providers promoted the program to their patients, it made them more likely to participate. Second, health coaches identified more patients through other population health activities such as annual wellness visits, transitional care management calls, and a program for self-monitoring blood pressure. Finally, as the program became more visible and familiar in small, rural communities, people were more likely to participate; some even inquired about participating in the program.

The three main barriers to enrollment included (1) the lack of clarity about eligibility requirements, (2) delays in implementing the health IT systems, and (3) patient preferences or perceptions about the benefits of preventive care. For example, at some implementing sites, provider champions said they were not clear on exactly what determined eligibility. One provider champion interviewed during the third program year guessed that she was under-referring, but said she didn't have a list of eligibility criteria to reference. Rural market managers corroborated this, describing how many PCPs "get stuck" only referring hypertensive and diabetic patients and have to be reminded to "think outside the box" to increase enrollment.

Delays in establishing the health IT systems also impeded the program's identification of eligible participants. Health coaches interviewed during the third program year still identified potential participants by reviewing patient lists in Excel spreadsheets, instead of using the automated process the awardee had envisioned.

Finally, staff believed that patient preferences and perceptions were a barrier to enrollment. For example, many residents in rural communities are not accustomed to preventive care and only seek medical services when they are sick. Respondents to both the clinician and staff surveys agreed that patients might have declined to enroll for the following reasons:

- The services or care seemed unnecessary or unimportant to the patient (70 percent of clinicians, 52 percent of non-clinicians)
- The patient did not understand the program or its potential benefits (48 percent of clinicians, 65 percent of non-clinicians)
- It was too much of a time commitment (46 percent of clinicians and non-clinicians)

In addition, the majority of clinicians who answered the survey thought patients might refuse to participate because they lacked the motivation to work on health goals. Half of the respondents to the staff survey thought patients might have refused because they were content with the old system of care.

2. Delivery of program services

a. Description of and changes to service delivery model

The Mercy ACO initiated health coaching in its urban practices in 2012 and began introducing the service in its rural practices one year later. Most rural hospitals (16 of the 25 hospitals participating in the HCIA R2–funded program) already had a health coach in 2013, but the health coach was considered a fairly new position and was not fully defined prior to the start of the program. The cooperative agreement standardized health coaches' responsibilities by

focusing their services around specific disease priorities and activities and by dedicating health coaching resources to patients most likely to benefit from the intervention. It also expanded the number of staff involved in health coaching and other population health activities at the rural hospitals and their affiliated clinics.

For example, rural market managers were hired as part of the HCIA R2–funded program to liaise between the Mercy ACO and the implementing sites and to standardize the program across the Central Iowa, North Iowa, and Siouxland regions. Similarly, the awardee appointed provider champions, who were practicing PCPs at participating sites, to attend Mercy ACO meetings, bring information from the ACO meetings back to their sites, and work with the health coaches to engage other clinicians and staff in the program.

The purpose of the health coaching component of the awardee's program was to (1) help participants manage their chronic diseases, (2) help participants set and achieve their self-defined health goals, (3) coordinate patient- and family-centered care services for participants, and (4) connect participants to community-based resources. The Mercy ACO adopted a short-term, intensive health coaching model, with interactions optimally taking place at least once a week for six weeks. The interactions were preferably conducted in person but could have taken place over the phone, especially for follow-up care after the initial face-to-face meeting. The components of the health coaching model were wide-ranging, but could have included (1) reviewing medical conditions and monitoring alert values, (2) reviewing medication lists, (3) providing self-management support and goal-setting, (4) closing gaps in care, (5) screening for depression and use of tobacco, (6) developing advance care directives, and (7) coordinating follow-up care. However, health coaches and rural market managers described these components more as guides than as strict protocol, and rural market managers encouraged health coaches to tailor services to meet participant needs. The duration of enrollment and the frequency of meetings varied depending on participants' levels of engagement and their progress meeting health coaching goals.

> "This [health coaching model] is a guideline, but the patients might be in another place. And then also, we want to give the coaches the opportunity to use their clinical judgment as well, and their critical thinking skills for what they feel may be best for that patient at that time."
>
> —Rural market manager

In addition, health coaches also handled annual wellness visits for Medicare beneficiaries, transitional care management after a hospitalization (follow-up with patients and coordination with a PCP within 14 days of discharge), and advance care planning (in-person meetings with patients to discuss and enact advance directives). During interviews in the third program year, the awardee said that as the program went on, there was more focus on documenting and billing for these services to help cover the costs of health coaching. Although these extra activities helped identify eligible health coaching patients and aligned with the organizational goal of providing value-based care, beneficiaries who only received these billable services were not considered direct program participants.

b. Evidence of service delivery effectiveness

Based on the available evidence, we have determined that Mercy ACO was partly successful in implementing the health coaching component of its program. The awardee's successes included hiring and training the health coaches and rural market managers and appointing the provider champions needed to implement the program at participating clinics in the three geographic regions. The awardee was also successful at increasing provider engagement over the three years of the cooperative agreement, although variations in the level of engagement persisted. Program staff believed the program was also successful in engaging program participants. However, the health coaches' activities varied across participating sites, with some sites deviating from the intervention's intended focus on short-term, intensive care management. The awardee also experienced delays and challenges in implementing the health IT component of the program, which delayed or otherwise hampered participant identification, provider engagement, and quality monitoring and improvement initiatives for health coaching.

Delivery of intervention services. Mercy ACO succeeded in providing health coaching services on schedule at the affiliated rural clinics of all 25 hospitals. Data on participants with at least one in-person encounter with a health coach from the start of the cooperative agreement through May 2017 (three months before the end of the initial agreement) suggest that:

- Each group of clinics associated with a single hospital enrolled, on average, 321 patients into health coaching.
- Each participant had an average of nearly four encounters with a health coach.
- The number of participants varied across sites affiliated with a single hospital. Most clinics co-located with a hospital served more than 100 participants, and most distant clinics served fewer than 20.

These data corroborate qualitative evidence that health coaches provided services primarily at the main hospital clinics because they had the largest patient volume. According to health coaches, enrollment levels at the distant clinics varied depending on how engaged the providers were and how accessible the clinic was.

Qualitative interviews and quantitative encounter data also suggest that the definition of the health coaching intervention varied across regions, with some sites deviating from the intended model of intensive care management. Interviews with program leaders and frontline staff indicated that health coaching was the primary focus of the intervention, but that over time the role of the health coach evolved to include other services, such as annual wellness visits for Medicare beneficiaries, transitional care management calls, and advance care planning. Health coaches also participated in a pilot program on self-monitoring of blood pressure and filled other gaps in care.

According to the regional market managers, population health staff resources varied by site, and health coach priority lists helped health coaches determine how to allocate their time. For example, some rural clinics were large enough to have a dedicated staff member conduct annual wellness visits. However, health coaches at smaller sites conducted all of the population health activities in addition to their coaching responsibilities. There was also variation in the type of health coaching services given in the three regions. The Siouxland region focused on annual

wellness visits and transitional care management calls; frontline staff in that region reported spending minimal time on traditional health coaching activities.

Encounter data reported by health coaches suggest that the majority of patients who were served had fewer than the six encounters recommended for the short-term, intensive health coaching intervention (Table II.2). However, it is important to note that these data include all activities logged by health coaches, and likely include some people who had annual wellness visits but did not receive traditional health coaching.

Table II.2. Number and percentage of enrollees by number of encounters with a health coach, by region

Number of encounters	Central (n = 3,189)		North (n = 3,207)		Siouxland (n = 2,090)	
	Number	Percent	Number	Percent	Number	Percent
1	1,203	37.7	1,077	33.6	802	38.4
2	572	17.9	596	18.6	378	18.1
3	334	10.5	355	11.1	394	18.9
4	227	7.1	259	8.1	188	9.0
5	159	5.0	210	6.5	109	5.2
6	111	3.5	129	4.0	57	2.7
7	91	2.9	104	3.2	35	1.7
8+	492	15.4	477	14.9	127	6.1

Source: CHIIC health coaching management system, September 21, 2017.

Note: Results are based on 8,496 health coaching enrollees. We identified health coaching enrollees by matching on first name, last name, and date of birth. We excluded participants with missing date of birth (375), participants who did not have at least one in-person visit (11,286), participants whose initial in-person visit occurred after May 31, 2017 (688), and participants with missing region (109). Because of a change in the system used to track participation, encounter data for the Central Region were only available through the ninth program quarter (November 2016).

Finally, the awardee failed to establish data connectivity between the Mercy ACO and some implementing sites. Even among successful sites, establishing data connectivity took longer than expected, and the delay negatively affected critical aspects of the health coaching program, including identifying participants, tracking care management tracking, monitoring participant outcomes, and working on quality improvement.

Through the 12th program quarter, the awardee continued working with sites to improve data connectivity, determine which data elements were still not flowing to the ACO, and develop a plan for data acquisition. In addition, the Central Region's transition from TAV Health to Datashop during the 9th program quarter disrupted care management tracking and made it difficult to assess health coaching activities and keep up-to-date counts of direct participants. Some health coaches resisted documenting their services in the care management software, and this may also have affected the quality of the encounter data.

Staffing and training. The awardee successfully hired and trained staff, reaching its three-year staffing projection of 57 full-time equivalents by the third program quarter.[11] The awardee also reported low staff turnover, especially during the third program year when the health coaching role had become more established and valued by clinicians. Eighty-four percent of non-clinicians and 73 percent of clinicians reported that they were very or somewhat satisfied with their role in the program.

The Mercy ACO encouraged all new health coaches and health coach assistants to attend an initial training as well as a refresher course after two years. The Mercy ACO exceeded its training goals in the first two years of the program, but fell short of its goals during the third program year. The awardee exceeded its initial targets because several rural market managers participated in the initial training session for health coaches, and more coaches than expected participated in the refresher training session. Nearly all (96 percent) respondents to the non-clinician staff survey reported that they received formal health coach training. The majority of respondents reported that they received informal training as well, including staff meetings (96 percent), team huddles (92 percent), mentoring (76 percent) and shadowing (80 percent), asking a colleague for help (96 percent), and self-study (87 percent). Most non-clinician staff strongly or somewhat agreed that the training helped improve their job performance.

Recruitment and engagement of providers. Provider engagement varied by site, but improved over the three years of the cooperative agreement. According to one provider champion, as PCPs began to realize the benefits of population health activities and to understand how health coaches could support them in caring for their patients, they were more likely to be receptive to the program. By the time the clinician survey was completed in spring 2017, nearly all clinicians who responded to the survey thought the program was very or somewhat effective (94 percent), reported that they had observed better patient care as a result of the program (91 percent), and thought the model should be set up in other clinical settings (90 percent). However, only about two-thirds of eligible clinicians responded to the survey.

Engagement of program participants. Frontline staff reported effectively engaging participants in the health coaching program. During interviews, health coaches described several successful strategies they used to engage participants in health coaching. These included offering visits after hours and on weekends, offering phone calls as an alternative to in-person visits, connecting participants with local resources, and showing participants how their efforts were paying off. Data collected through program surveys support this finding. For example, 76 percent of non-clinicians and 81 percent of clinicians strongly or somewhat agreed that the program successfully engaged participants. Similarly, among clinicians who responded to the survey, the most common reason given for the program achieving its goals was successful participant engagement (55 percent of respondents). Eighty-two percent of non-clinician staff survey respondents believed the program had a positive impact on participant engagement. However, frontline staff acknowledged that not all participants were engaged, and that the onus was ultimately on the participants to be responsible for their health behaviors; participants at times refused services or disengaged if they were non-symptomatic.

[11] There was one exception: one participating hospital and its affiliated clinic were without a health coach for about half of the second program year.

c. Barriers and facilitators associated with service delivery effectiveness

Participating sites' membership in the Mercy ACO, effective communication strategies, and the rural sites' small size facilitated implementation across the hospitals and their affiliated clinics. First, the Mercy ACO had a commitment to population health activities at its highest levels of leadership, which provided a foundation for implementing the program and engaging sites. Second, staff at all levels of the program indicated that effective and constant communication between clinic staff, regional management, and program leaders was essential for keeping staff informed and increasing engagement over the three years of the cooperative agreement. Communication strategies included (but were not limited to) monthly Mercy ACO meetings, regular meetings between rural market managers and office managers, regular health coaching meetings, regular provider champion meetings, internal clinic huddles, and site visits conducted by the rural market managers. Lastly, program leaders reported that the participating rural sites' small physical size and staff size made it easier to raise the program's visibility and integrate its services compared with the urban sites.

Major barriers to program implementation included competing priorities in resource-constrained clinics, the wide geographic reach of the program, and problems implementing the health IT component. First, in Year 1, health coaches were pulled from their health coaching work to fill other administrative and clinical gaps. To minimize this, awardee leaders defined the health coaches' roles and responsibilities more clearly, and conducted site visits to educate clinic staff about the health coach's role. Frontline staff thought the health coach's role was more clearly defined and integrated into the clinic workflow during the second and third program year. As of the third site visit, some provider champions still reported that lack of time and fast-paced workflows were a barrier to engagement.

> "Because as providers we have so many other things on our plate, … I feel like if we're not constantly hearing what the program is working on, what we're doing with the program, what successes we have, what challenges we have, [it's] kind of off our radar to be honest."
>
> —Provider champion

Second, the program's wide geographic expanse and the large number of implementing sites across the three regions of the Mercy Health Network challenged implementation. Information presented at monthly ACO meetings did not always trickle down to staff at the implementing sites, and staff from implementing sites found it difficult to attend ACO meetings in person. Furthermore, the awardee depended on local staff to champion the program, because some rural sites thought the program was coming from and benefiting the ACO in Des Moines.

Third, Health IT was a barrier to program implementation at many sites, depending on each site's connectivity and ability to use data for its intended purposes. By the end of the program, the awardee still struggled to connect some sites' EMR systems to the disease registry because many of the implementation sites used different EMR platforms. Some frontline staff were able to use performance data from the ACO to get providers engaged in the program, but other frontline staff reported that the data their practice received from the ACO were not reliable enough to show the PCPs or use for quality improvement. Awardee leaders reflected that health IT development was much more challenging than they anticipated, and they wished they had managed expectations better to avoid the negative impact on site engagement.

Finally, external payment policies affected site engagement in several ways. On the one hand, awardee leaders believe that the release of the Medicaid Access and CHIP Reauthorization Act of 2015 (MACRA) regulation was well timed and promoted engagement at implementing sites during the third program year. MACRA established the Quality Payment Program, through which clinicians are rewarded for delivering high-quality care and improving performance.[12] These new payment policies, which will affect reimbursement levels at many participating sites, validated the importance of population health activities and encouraged sites to take advantage of the health coach role. Furthermore, participating sites turned to the Mercy ACO for help interpreting MACRA regulations and implementing processes to maximize reimbursement. Awardee leaders believe that MACRA gave participating practices an incentive to work together and share best practices for population health improvement.

On the other hand, some of the participating sites were designated rural health clinics (RHCs), which are not affected by MACRA payment adjustments and therefore not incentivized by MACRA to deliver high quality care. Similarly, these RHCs are paid an all-inclusive rate for services provided to a patient in a single day. Therefore, unlike non-RHC sites, they cannot use annual wellness visits to offset the cost of the health coach, unless the patient is willing to come back to the clinic on a separate day.

C. Assessment of perceived program effects on the delivery of care and outcomes

Survey data and qualitative interviews suggest there was a strong consensus among frontline staff and awardee leaders that the program had a positive impact on patient care. For example, more than three-quarters of non-clinician staff (83 percent) and clinicians (77 percent) believed their ability to provide care was better as a result of the program, and nearly all non-clinician staff (90 percent) and clinicians (94 percent) believed the program was either very or somewhat effective at achieving its goals. At least two-thirds of all survey respondents believed the program had positive impacts on a variety of aspects of care (Table II.3).

[12] For additional information on the Quality Payment Program, see the Quality Payment Program website, available at https://qpp.cms.gov/.

Table II.3. Number and percentage of clinicians and non-clinician staff that believed the program had a positive impact on aspects of care

Aspects of care	Clinicians (%)	Non-clinician staff (%)
The quality of care and services you provide to patients	89 (85%)	42 (84%)
Your ability to provide care or services that are responsive to participants' preferences, needs and values	77 (73%)	42 (84%)
Access to care or services for all participants	71 (67%)	37 (74%)
Achievement of participants' health goals	89 (84%)	43 (86%)
Participant satisfaction	84 (79%)	41 (82%)
Participant quality of life	80 (75%)	41 (82%)
Care coordination	92 (87%)	43 (86%)
Patient self-management	80 (76%)	40 (80%)
Patient education	94 (89%)	41 (84%)
Patient engagement	82 (82%)	41 (82%)
Patient use of community resources	71 (68%)	39 (78%)

Source: HCIA R2 evaluation survey of participating non-clinician staff (fall 2016), and clinician staff (fall 2017).

Note: Because respondents can choose more than one aspect of care as improved, percentages sum to more than 100 percent. The clinician survey had a response rate of 64 percent with a sample of 230 potential respondents; the non-clinician staff survey had a response rate of 96 percent with a sample of 96 potential respondents.

Information gathered during interviews in the third program year corroborated these findings. There was consensus among the frontline staff that the program goals were attainable and aligned with the health coaching activities, although some frontline staff thought it would take time for patient outcomes to become observable. For example, some frontline staff believed the program reduced the number of ED visits because participants called health coaches with questions about their symptoms instead of going to the ED. Frontline staff believed the program improved care delivery because chart reviews, health coaching visits, and annual wellness visits identified patients who were due for preventive services. Health coaches provided anecdotal evidence of health coaches recommending cancer screenings that led to the early detection of cancer in several participants.

> "I think one of our biggest barriers to [making] the program successful, might just be staffing ... I know that's a challenge everywhere, but if you wanted to see great improvement in those numbers I feel like we would need more health coaches, or more staff to utilize coaches appropriately … I think we have a lot of challenges – more so in rural America than elsewhere."
>
> —Provider champion

Non-clinician staff believed that buy-in to the program from clinicians, support from program leaders, and engagement on the part of the participants helped the program be effective. Similarly, in interviews during the third program year, regional and frontline staff attributed the program's effectiveness to communication, collaboration, and engagement of staff across the network, as well as having a dedicated health coach position at the clinics to focus exclusively on population health.

Several respondents said that if the program were found to be ineffective, they believe that would be due to (1) participant characteristics beyond their control, such as poverty or lack of

compliance with the program; (2) three years not being long enough to observe outcomes; (3) providers not having enough time for significant redesign of processes; (4) health coaches not having the capacity to serve all of those who qualify; and (5) Mercy ACO not being able to provide performance data in real time so sites could monitor implementation and make more informed improvements to the program.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Given the large number of implementing sites and program staff and the degree of flexibility clinics were given in providing health coaching services, we expect the program's impact to vary across clinics. Three aspects of program implementation with implications for measuring changes in patient outcomes and costs are:

- **Program enrollment.** Although the awardee enrolled enough people to evaluate the effects of the program, frontline staff reported enrolling patients with health care goals and needs (such as smoking cessation, weight loss, and chronic condition risk factors) who may benefit less from the short-term, intensive health coaching intervention than the original target population of patients with more than one chronic condition. The broad eligibility criteria will likely weaken the program's overall effects on health outcomes and costs.
- **Timing of implementation.** All staff positions were filled by the third program quarter. As a result, we might expect to see program effects as early as the end of the first program year. However, qualitative evidence suggests that program operations improved over time, and the health coaches were better utilized within the clinics.
- **Care delivery.** Based on our knowledge of the health coaching activities, we might expect to find higher rates of preventive services (such as screenings, immunizations, and primary care visits) and more control of hypertension and diabetes among participants enrolled in the health coaching program. However, the lack of standardization of the health coaching program and deviation from the short-term intensive care management model might dilute the impact we expect to see on any particular health outcome. For example, because health coaches in the Siouxland Region focused more on annual wellness visits than on traditional health coaching activities, we might expect to see higher rates of preventive services, but less impact on diabetes and hypertension control.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter updates our previous findings on the baseline characteristics of the treatment group as described in the second annual report. In that report, due to a lack of data, we had assumed we would have to evaluate the impact of the practice-level component on outcomes and, as a result, presented information based on all Medicare beneficiaries who sought care at a participating clinic. During the past year, however, we received a copy of the awardee's health coaching encounter-level database and were able to use it to identify Medicare beneficiaries enrolled in the short-term, intensive care management intervention—the component of the awardee's program that is expected to have the largest impact on costs and health outcomes. Armed with these data, we have shifted the focus of our evaluation from the (weaker) indirect practice-transformation component to the (stronger) direct health coaching component. The results in this chapter now reflect the baseline characteristics of Medicare beneficiaries who actually had an in-person visit with a health coach.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Catholic Health Initiatives

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	11,982[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	3,484
Likelihood of all-cause hospitalizations	8,430
MDE sample size requirement to detect 20% effect	
Total expenditures	871
Likelihood of all-cause hospitalizations	2,108
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline. At some clinics, intervention was expansion of existing program.

Table III.1 *(continued)*

Evaluability domain	Response
Claims sufficient to identify treatment and comparable comparison group?	Yes, study population identified through attribution to participating or comparison clinics
Likelihood of solid comparison group	No serious issues; proceeding with comparison group selection
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician, staff, and beneficiary surveys
Implementation data that will be analyzed	Encounter data used to measure number of encounters and duration of enrollment

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We plan to conduct a rigorous impact analysis of the Catholic Health Initiatives' health coaching intervention. (We will not explicitly measure the impact of the program's lean process improvement component at the practice level.) The treatment group will consist of all beneficiaries who sought care at a participating clinic, including those who did not enroll in health coaching. The comparison group will consist of beneficiaries who are patients at CAH-affiliated clinics in Iowa and Nebraska that are not affiliated with Catholic Health Initiatives. We will match at the beneficiary level.

The analysis will need to distinguish between those clinics that had mature coaching programs at the time of the HCIA R2 cooperative agreement and those that had not. Therefore, we will assess the intensity of health coaching programs at each clinic in consultation with the implementation team. We will exclude clinics with fully developed programs prior to award from the impact analysis.

B. Characteristics of Medicare and Medicaid participants at baseline

As noted, Catholic Health Initiatives is implementing two overlapping interventions. One is a practice-level component focused on improving both population health and processes of care. All patients who seek care at participating clinics potentially benefit from this practice transformation component and are considered indirect program participants. The other is a participant-level component that involves the direct provision of care management services through trained health coaches who are embedded in the clinics, usually on a part-time basis.

Although all patients are eligible to receive health coaching services, the awardee initially encouraged participating clinics to target those with one or more chronic condition specified by the awardee. Patients were actively recruited into health coaching, in contrast with the practice transformation component. Those patients who agreed to meet with a health coach were enrolled in the program; they are considered direct program participants. The results in this section are based on Medicare FFS beneficiaries who had at least one in-person encounter with a health coach.

According to self-reported data, the awardee enrolled over 13,000 patients (10 times higher than its original three-year target of 1,295) into health coaching during the full three years of the program[13]. However, the awardee was unable to provide Medicare or Medicaid identification numbers for those who enrolled in health coaching. As a result, we identified them in Medicare enrollment files by using their name, date of birth, gender, and state of residence as reported in the quarterly encounter-level database—a less exact approach than using identification numbers, but one that led to a fairly high match rate nonetheless, likely due in large part to the less densely populated rural catchment areas targeted by the program.

In presenting baseline characteristics for direct participants in the health coaching component, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when they became eligible for awardee-provided services (that is, their enrollment date) and who had met all evaluation criteria for a period of at least 90 days within the baseline year (the 365 days immediately before their enrollment). In addition, they had to enroll in the awardee's program on or before May 31, 2017, to ensure a long enough run-out period to capture nearly all claims for the most recent participants.

As noted, we identified 8,595 unique enrollees from our analysis of the encounter-level database submitted by the awardee. This is based on those with at least one face-to-face encounter with a health coach, the criterion used by the awardee to define enrollment. Of these, we were able to link 5,634 to the Medicare enrollment data base using name, date of birth, gender, and state of residence. Of these, 4,678 met our inclusion criteria for the impact evaluation and are included in the analysis of baseline characteristics. Most of the 956 linked Medicare beneficiaries we excluded from the analysis were enrolled in Medicare Advantage (414) or were not enrolled in Part B (358).

There are two caveats to this analysis. First, because we were unable to identify participants who only received an annual wellness visit from a health coach (reported in the encounter data base), our analysis includes some beneficiaries who were likely in better health than those enrolled in the short-term, intensive care management intervention. Second, our analysis does not include enrollees from the Central Region during the final nine months of the program. During those nine months, clinics there were unable to provide encounter data to the awardee.

Table III.2 shows beneficiaries' baseline demographic and health status characteristics. Almost half of the health coaching participants were 75 or older; only 10 percent were younger

[13] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from participant encounter data furnished by the awardee to us. The number of participants will differ for several reasons. First, the awardee submitted these data to us prior to August 31, 2017, so it does not include all participants. The encounter data sent by the awardee is missing records from the Central region for most of the third program year (because that region started using a new care management software); however, these records from the Central region are included in the direct participant count reported to the implementation and monitoring contractor. In addition, we restricted enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS and those participants for whom the encounter data has valid indentifiers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least 6 months of post-enrollment data.

than 65. Almost all of the participants were white (98 percent) and three out of five (60 percent) were female.

For most participants (82 percent), age was the original reason for their Medicare eligibility. Eighteen percent of participants were dually eligible for Medicare and Medicaid. The average hierarchical condition categories (HCC) risk score for health coaching participants (1.38) was nearly 40 percent higher than the average for Medicare FFS beneficiaries nationwide. Almost half of the health coaching participants had HCC risk scores that were lower than the national average, however. This could reflect the range of risk factors and health needs that providers used to identify participants. The below-average HCC scores could also reflect the inclusion in the encounter data of beneficiaries who only had a wellness visit with a health coach, but not enrolled in the health coaching program. Nonetheless, given the focus of the program, it is not surprising that participants of the health coaching intervention, who were recruited into the intervention based on need, averaged relatively high HCC risk scores.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Catholic Health Initiatives' health coaching program through May 31, 2017

	All participants (n = 4,678)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	477	10
65 to 74	1,930	41
75 to 84	1,509	32
85 and older	762	16
Gender		
Female	2,821	60
Male	1,857	40
Race		
White	4,581	98
Black	8	0.17
American Indian, Alaska Native, Asian/Pacific Island American, or other	48	1
Hispanic	6	0.13
Original reason for Medicare eligibility		
Old age and survivor's insurance	3,815	82
Disability insurance benefits	848	18
End-stage renal disease (ESRD)[a]	15	0.32
Hospice[b]	4	0.09
Medicare/Medicaid dual status, percentage dual[b]	700	15
HCC score[c]		Statistic
Mean		1.36
25th percentile		0.54
Median		0.95
75th percentile		1.72

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2017.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date of a participant's first face-to-face encounter with a health coach. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Medicare beneficiaries enrolled in the short-term, intensive care management intervention also had relatively high rates of service use and Medicare expenditures, on average, in the 365 days before enrollment. Table III.3 shows baseline utilization and expenditure data for a common set of measures. We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments in total and by major types of services.

The total average PBPM Medicare payment during the baseline year was $1,071, roughly 35 percent above the 2014 national average of $792 for all Medicare FFS beneficiaries.[14] Average PBPM Medicare payments for acute inpatient ($380), outpatient ($319), and physician ($161) services were the largest drivers of the total cost of care. The rate of acute care hospitalizations was 438 per 1,000 Medicare FFS beneficiaries per year during the baseline year, again far above the national average of 274 per 1,000 Medicare FFS beneficiaries in 2014. The rate of outpatient ED visits per year during the baseline year was 781 per 1,000 Medicare FFS beneficiaries.

The rate of specialist visits in any setting per year was 6,508 per 1,000 Medicare FFS beneficiaries during the baseline year, whereas the rate of specialist visits in ambulatory care settings per year was 5,409 per 1,000 Medicare FFS beneficiaries. The rate of primary care visits in any setting per year was 3,574 per 1,000 Medicare FFS beneficiaries, whereas the rate of primary care visits in ambulatory settings per year was 2,391 per 1,000 Medicare FFS beneficiaries.

Somewhat surprisingly, health care expenditures and service use in all major settings in the quarter before enrollment were significantly higher than the rates of expenditures and utilization for the entire year. They were even significantly higher in the quarter immediately before enrollment than they were during the next most recent quarter. For example, total PBPM expenditures more than doubled in quarter 4 of the baseline year ($1,794) compared with quarter 3 of the baseline year ($808). This finding indicates that participants were likely to be recruited (or were more likely to enroll) in the health coaching program when they were experiencing a sudden deterioration in health and an increase in health care needs. This finding is consistent with our understanding of program recruitment; participating clinics reviewed hospital discharge and ED visit data to identify beneficiaries who would benefit most from health coaching.

Through the combined impact of both its health coaching and practice transformation components, Catholic Health Initiatives expected to reduce preventable hospitalizations by 12 percent and reduce use of the ED for non-emergencies by 30 percent. The awardee also expected to increase the use of primary care services by 30 percent. If the reductions in preventable hospitalizations and ED use were realized, the awardee anticipated a reduction in the total cost of care for participating rural communities by 2.08 percent for Medicare beneficiaries and 1.92 percent for dually eligible beneficiaries by the end of August 2017.

[14] The mean expenditure was drawn from the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2015. Hospitalization rates were taken from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Catholic Health Initiatives' health coaching program through May 31, 2017

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	4,678	4,498	4,581	4,675	4,640
Average Medicare expenditures PBPM[a]					
Total	1,071 (31)	860 (52)	815 (43)	808 (35)	1,794 (66)
Acute inpatient	380 (18)	271 (39)	246 (27)	236 (20)	760 (35)
Inpatient other[b]	17 (4)	15 (7)	19 (9)	16 (8)	17 (5)
Outpatient[c]	319 (7)	298 (10)	285 (10)	280 (12)	412 (12)
Physician services	161 (5)	143 (6)	143 (5)	149 (6)	207 (8)
Home health	19 (1)	14 (2)	16 (2)	19 (2)	27 (3)
Skilled nursing facility	143 (10)	85 (15)	72 (10)	76 (10)	336 (31)
Hospice	2 (1)	2 (1)	2 (1)	2 (1)	3 (2)
Durable medical equipment	31 (3)	30 (3)	32 (3)	30 (3)	32 (4)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	438 (14)	275 (18)	246 (17)	289 (27)	931 (32)
Outpatient ED visits[e]	781 (25)	598 (30)	593 (30)	657 (38)	1,270 (46)
Primary care visits in any setting	3,574 (96)	3,075 (115)	3,106 (111)	3,132 (115)	4,970 (157)
Primary care visits in ambulatory settings	2,391 (67)	2,249 (79)	2,331 (78)	2,270 (82)	2,715 (85)
Specialist visits in any setting	6,508 (139)	6,150 (177)	6,323 (207)	6,062 (157)	7,523 (196)
Specialist visits in ambulatory settings	5,409 (104)	5,388 (130)	5,411 (142)	5,314 (130)	5,557 (127)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	28 (1)	6 (<0.5)	5 (<0.5)	6 (<0.5)	19 (1)
Percentage with an outpatient ED visit[e]	37 (1)	11 (<0.5)	11 (<0.5)	12 (<0.5)	22 (1)
Percentage with a 30-day readmission among all discharges	14 (1)	10 (2)	17 (2)	13 (2)	16 (2)
Percentage of participants with a readmission among all participants	3 (<0.5)	1 (<0.5)	1 (<0.5)	1 (<0.5)	1 (<0.5)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2017.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

The awardee's payment model is based on the creation of a rural MSSP ACO for Mercy CAHs and their affiliated clinics. The awardee would supplement the new shared savings arrangement with CMS with FFS billing for specific health coaching services, including transitional care management, advance care planning, and annual wellness visits. By the end of the three-year cooperative agreement, all of the rural hospitals participating in the health coaching program had joined the rural MSSP ACO.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Catholic Health Initiative's payment model relies primarily on shared savings supplemented by FFS billing for specific services conducted by health coaches, including transitional care management, advance care planning, and annual wellness visits. Originally, the awardee expected the network CAHs and their affiliated clinics to participate in the existing Mercy ACO. The savings generated by the health coaching intervention would be used to cover the cost of the intervention. However, expanding the health coaching program to rural providers failed to generate the expected savings. The awardee attributed this to the unique cost structure and reimbursement system for CAHs.

Within a shared savings contract, providers have a direct financial incentive to eliminate unnecessary use of medical services; providers who participate in an MSSP ACO have the possibility of receiving a percentage of any savings generated by a decrease in average costs per attributed beneficiary. However, the vast majority of costs incurred by CAHs are fixed. As a result, a reduction in the number of ED visits or hospital admissions can actually lead to an increase in average costs per patient. And given the cost-based reimbursement system Medicare uses to reimburse CAHs, an increase in allowable costs leads to an equivalent increase in Medicare payments. Therefore, the inclusion of CAHs and their affiliated clinics within the existing MSSP ACO can actually make it more difficult for an existing ACO to lower the average costs across all of its attributed beneficiaries and to generate shared savings for its providers.

In January 2017, after realizing the perverse effect that including CAHs in its existing MSSP ACO could have on average costs per attributed beneficiary, the awardee decided to establish a new and separate rural MSSP ACO for its rural hospitals and their affiliated clinics. The new MSSP ACO includes all 23 CAHs and their affiliated clinics that participated in the health coaching program. To participate in the rural MSSP ACO, CAHs must pay an annual fee to support the infrastructure and employ at least one health coach. The Mercy ACO will continue to provide administrative services to support the Medicare shared savings program, including provider credentialing, compliance and reporting, and system maintenance. In the event that

there are shared savings, the awardee uses a distribution formula based on the programs that the providers participate in.

In addition to the potential for achieving shared savings, the awardee hopes the rural MSSP ACO will create a mechanism to drive operational processes to improve clinical integration and deliver coordinated care across a complex network of providers. The advantages of participating in the rural MSSP ACO also include efficient reporting of measures for the Physician Quality Reporting System, favorable Merit-based Incentive Payment System performance through clinical integration, and sharing of claims data. The awardee acknowledged that high fixed costs will make it difficult for CAHs to achieve savings even in the new rural MSSP ACO, but the other benefits of joining a rural ACO will still make the effort worthwhile.

The awardee is also pursuing opportunities to cover the cost of health coaching through FFS billing for population health activities. Health coaches can bill for time spent providing transitional care management, advance care planning, and annual wellness visits. In addition, the participating rural clinics are identifying opportunities to deliver appropriate preventive services—such as hemoglobin A1c tests, mammography screenings, and colorectal cancer screenings—that help improve patient outcomes, reduce costs in the long run, and can be billed under traditional FFS.

C. Status of the payment model

All 23 CAHs and their owned or affiliated clinics participating in the health coaching program have joined Mercy's new rural MSSP ACO. The awardee has received positive feedback from providers who view participation in the MSSP ACO as a way to position themselves for future value-based payment reforms in the state. The awardee is continuing to explore opportunities for FFS billing by health coaches.

D. Factors associated with the development of the payment model

The awardee said several factors affected development of its payment model. Facilitating factors included strong leadership engagement and organizational focus on alternative payment models and population health. Challenges faced by the awardee involved the characteristics of rural health providers and the rural patient population. Also, because of the small size of the rural patient population and low rates of private health insurance coverage, commercial payers have shown little interest in pursuing value-based contracts to support the program. However, the Mercy ACO hopes to expand its agreements with commercial payers focused on urban providers and to try to include the rural population over time.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The Mercy ACO was making continued progress on its sustainability efforts, most notably on encouraging site participation in the rural MSSP ACO and engaging hospital and clinic leaders. CMS also granted Mercy a six-month no-cost extension. The extension will allow the awardee to continue developing the disease registry, which the Mercy ACO uses to track clinical data for its attributed population and which is used by the health coaches and clinic staff to identify eligible health coaching patients and performance improvement projects. The awardee did not plan to scale the program. However, the health coaching program was implemented at all new sites that joined the Mercy ACO during the cooperative agreement and will be implemented at new sites going forward. Finally, although the awardee did not report any plans to replicate the program, it had received inquiries from other rural health systems interested in implementing similar programs.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, the awardee reported a number of strategies to sustain its program, including leveraging sites' participation in the Mercy ACO, gaining buy-in from

organizational leaders, requiring sites to gradually assume program costs, and committing to sustain the positions of regional program staff.

First, all implementing sites agreed to join the Mercy Rural MSSP, which requires sites to engage in population health activities. Second, the awardee began working to gain buy-in from organizational leaders at participating sites by highlighting the program's long-term cost savings and moral obligation to give its patients the highest quality of service. Third, the awardee required sites to assume more of the program's cost with each program year to reduce their dependency on funding from the cooperative agreement. Sites assumed 10 percent of the costs in Year 1, 20 percent in Year 2, and 50 percent in Year 3. They will be expected to cover the full health coach salary at the end of the cooperative agreement. Finally, the awardee planned to sustain the positions of rural market managers after the end of the cooperative agreement.

C. Implementing the SSR plan: progress and changes

Sustainability. As of the third program year, Mercy ACO continued making progress on the sustainability efforts it had initiated the first two program years. All implementing sites had joined the Mercy Rural MSSP and in doing so, committed to improving population health, including sustaining the health coaching position. Moreover, the awardee received a no-cost extension through February 2018, which the awardee will use to ensure structures are in place so that the sites can continue participating in the ACO's population health activities beyond the cooperative agreement. Specifically, the awardee reported plans to continue developing the disease registry so that it can effectively identify patients, document care management activities, and report outcomes. The Mercy ACO also reported plans to use the no-cost extension to continue engaging administrative leaders at critical access hospitals on the value of the health coach teams. The awardee reported that the disease registry will help engage leaders by providing "additional evidence of the efficacy of our population health efforts, including the work of the health coach teams."

As planned, sites paid for 50 percent of program costs in the third program year and were reportedly set to fund the positions after the end of the award period. The sustainability of the quality improvement component of the program varied by region, with the quality improvement staff position eliminated in the Central region, and the funding for the position divided between eight local staff members in the Northern region. The role will likely be sustained in the Siouxland region, because it existed before the cooperative agreement began.

Scalability. Mercy ACO did not plan to formally scale the program, although the health coaching program was implemented at all sites that joined the Mercy ACO during the cooperative agreement, and will be implemented at any sites that join the ACO going forward.

Replicability. Mercy ACO did not report formal plans to replicate the program, although it received inquiries from other health systems interested in implementing similar programs.

D. Factors associated with progress toward implementing the SSR plan

Awardee leaders said that (1) support at the highest level of Mercy leadership, (2) the sites' existing membership in the Mercy Health Network, and (3) MACRA legislation all made it possible to sustain the program at implementing sites.

One awardee leader explained that population health had been a priority for the network before the cooperative agreement. "Our highest-level executive always said that if you're part of Mercy Health Network, [population health management] is part of the expectation. Commitment from the highest level of leadership has been a huge reason why this work is sustainable." Moreover, awardee leaders believed that MACRA facilitated buy-in and engagement from all levels of the organization because it emphasized the need to focus on value-based care. "MACRA regulations beefed up our case for why we needed this work. It keeps people engaged," one respondent explained.

Despite leadership engagement and MACRA's facilitation of sustainment, the awardee recognized that the Mercy ACO contracts would end in just three years, and rural administrators would need more financial and clinical data to continue supporting the program in the long term. The awardee reported using a variety of strategies to keep leaders engaged after the cooperative agreement ends.

First, the awardee hopes that the disease registry and care management platform will be able to track and report program outcomes that can be shared with administrative leaders. Moreover, the awardee leaders created a document to show the potential positive or negative adjustments to reimbursement as a result of MACRA, which can help sites visualize the program's potential impacts on reimbursement. Third, rural market managers will continue directing the work of health coaches, specifically helping managers and administrative leaders at the implementing sites to understand how to support health coaches. Health coaches who work within local clinics will be trained to engage leaders and other stakeholders in their sites by promoting their services, sharing data and progress reports with providers and clinic administrators, and constantly improving their processes.

Finally, the awardee reported a key challenge for designated RHCs that it could not overcome without an alternative payment model. Designated RHCs, a classification that pertains to some participating sites, are reimbursed at an all-inclusive rate. As a result, these sites cannot be reimbursed for annual wellness visits provided on the same day that a patient is already in the clinic for another reason, and these sites will not benefit directly from MACRA payment adjustments. Sites with an RHC status have been less engaged than other sites, and the inability to bill for annual wellness visits will be a threat to financially sustaining the health coach's role.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insight into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Catholic Health Initiatives' health coaching program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicare beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: The Children's Home Society of Florida

April 18, 2018

Margaret H. Coit (Mathematica Policy Research)
Genna R. Cohen (Mathematica Policy Research)
Marisa Morrison (RTI International)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Children's Home Society of Florida, a nonprofit advocacy organization committed to protecting and supporting children and families, used funding from HCIA R2 to support the Evans Community School and Pine Hills community by implementing a program with two primary components: (1) patient navigation and (2) medical, dental, and behavioral health care

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

services (Table I.1). The awardee provided patient navigation services at two locations: (1) the Evans Community School, in an area of the main building known as the Hub, and (2) the Pine Hills Wellness Project office, a mile away from the school. Patient navigators at the Hub saw students and their families as well as community members on campus and connected them to a variety of health and social services, including housing supports, employment supports, food pantries, child care, and services that connect individuals to health insurance. Patient navigators who work out of the Pine Hills Wellness Project office did outreach throughout the community, connecting individuals to the same types of services as did the navigators at the Hub. The awardee also provided health care services at these two locations. Children's Home Society employees provided behavioral health services to students on the Evans Community School campus at the Hub. The awardee's Federally Qualified Health Center (FQHC) clinical partners[3] provided medical and dental services to students and community members at the Evans Health and Wellness Center, also on the Evans Community School campus.[4] Finally, the awardee used HCIA R2 funding to sponsor an optional, weekly after-school program called the Student Ambassadors to provide health education and student-to-student peer engagement to Evans Community School students. However, HCIA R2 is one of several funding streams that supported the activities offered through the Evans Community School and the Evans Health and Wellness Center.[5] As such, many participants likely received services that were paid for by a source or sources other than HCIA R2 funding. In addition, because clinical and administrative staff working at the Evans Health and Wellness Center were employees of the clinical partner and were not funded by HCIA R2, they were not considered program staff.

The origins of the Evans Health and Wellness Center (previously known as the Evans Wellness Cottage) predate HCIA R2. In 2012, Children's Home Society and its partners transformed the Evans High School into the Evans Community School. The community school model seeks to bring together the school and other community resources to promote academic achievement, strong families, and a healthy community. About the same time, Children's Home Society secured a grant through the Affordable Care Act to build a school-based Federally Qualified Health Center (FQHC), the Evans Wellness Cottage (later renamed the Evans Health and Wellness Center), on the Evans Community School campus. In addition, a community-based patient navigation pilot program previously operated outside of the school. Through its HCIA R2 cooperative agreement, Children's Home Society built on the existing community school infrastructure and its partnerships to expand the previous patient navigation pilot program to the school and the entire Pine Hills community and to offer behavioral health services at the Hub on the Evans Community School campus.

[3] True Health served as the awardee's clinical partner from October 2014 (the program's launch) to December 2016. Orange Blossom Family Health served as the awardee's clinical partner from January 2017 to August 2017.

[4] For the first four months of the cooperative agreement, the FQHC clinical partner provided health services to students out of the Hub, while the physical space for the FQHC was under construction.

[5] Although not funded by HCIA R2, the Evans Health and Wellness Center, now operated by Orange Blossom Family Health, is a freestanding mobile home at the back of the Evans Community School campus. Central to the community school model, the Center is designed to interact directly with the patient navigation and behavioral health services supported by HCIA R2.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The Children's Home Society of Florida (CHS) received HCIA R2 funding to implement the Evans Health and Wellness Center and the Pine Hills Wellness Project, a multipronged program intended to improve access to health care for individuals living in Pine Hills, FL.
Major innovation	Community Schools are a nationally recognized innovation to integrate community partners and provide comprehensive services that promote health and well-being. The awardee developed a program within the Community School model to expand access to health-related services to the students and others affiliated with the school and to the community at large.
Program components	Patient navigation Medical, dental, and behavioral health care services
Target population	All residents of Pine Hills, FL
Theory of change/ theory of action	CHS hypothesizes that implementing a program that incorporates patient navigation and direct services for the Pine Hills community will lead to lower cost of care, better use of appropriate services, and enhanced patient outcomes.
Payment model	Value-based payments, shared savings, partial or full capitation for medical services
Award amount	$2,078,295
Effective launch date[a]	10/1/2014
Program setting	Evans Community School
Market area	Urban
Market location	Zip code 32808 and 32818 in Pine Hills, FL
Target outcomes	• Decrease in emergency department (ED) visits per beneficiary • Increase in percentage of Medicaid/CHIP population receiving timely health care • Increase in participants' experience with care • Decrease in percentage of female participants younger than 18 who are pregnant • Decrease in percentage of youth enrolled at Evans Community School who report risky health behaviors • Decrease in overall cost of care • Decrease in percentage of participants with asthma who have one or more emergency department visits

[a]After the initial planning period, the awardee's program became operational as of this date.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we conclude that the awardee was partly successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on six factors. First, Children's Home Society enrolled 6,017 participants—80 percent of its revised enrollment target—by the end of the three-year cooperative agreement. Second, the awardee delivered services as intended for most participants, with the exception of the time during and after it changed clinical partners, when it temporarily halted medical services and curtailed dental services indefinitely. Third, Children's Home Society experienced turnover among community-based navigation staff and in one awardee leadership position; however, turnover affected recruitment efforts more than service delivery. Fourth, whereas Children's Home Society was

mostly successful in engaging external community providers and organizations, it struggled to attain the level of engagement that it desired with its initial clinical partner. Fifth, Children's Home Society was mostly successful in engaging student participants—especially through the Student Ambassadors program—but staff reported continued barriers connecting with and maintaining trust among community residents. Sixth, staff who completed the survey felt that the program was making a difference in meeting the needs of the community and had positively impacted participants' health goals.

Impact evaluation. Due to the lack of timely Medicaid data by the time of the final report, we do not anticipate being able to conduct a rigorous impact analysis for Children's Home Society.

Payment model. The awardee has conceptualized a payment model that would use a per beneficiary per month (PBPM) fee from Medicaid managed care organizations (MCOs) for primary and preventive medical, dental, outpatient behavioral health care, and community health and wellness promotion activities. Although the awardee has not begun any negotiations with payers, they plan to continue talks with local hospitals in order to tell them about the services they have to offer and explore potential partnerships.

Sustainability plans. Despite experiencing challenges finding funding for its program after the cooperative agreement ended, the Children's Home Society reported that it will sustain some aspects of the program, including a similar program that was newly implemented in the third program year in a nearby neighborhood. The awardee reported that challenges obtaining data prevented it from being able to prove the program's value to engage payers.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of non-clinician staff on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July to October 2016 with nine administrative and health navigation staff members and achieved a response rate of 67 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that Children's Home Society was partly successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

Children's Home Society recruited and enrolled two types of participants: (1) Evans Community School students and their families and (2) Pine Hills community residents. Although the awardee used different strategies to recruit each participant type, they used the same enrollment strategy for students and their families and for community members. Because HCIA R2 funding was only one of many funding streams supporting the awardee's efforts, the awardee considered all participants as indirect participants.

To recruit Evans Community School students and their families, school-based navigation staff operating out of the Hub informed students and families of available services, referred students and their families to the program, and assisted them in accessing care (for example,

making sure that students and family members had appropriate insurance, students had their parents sign consent forms, and students notified their teachers of appointment-related absences). Family members were most commonly recruited through referrals to the program from their student children. For example, if a student indicated to a member of staff that a parent had lost a job, staff might advise the student to refer the parent to the Hub to receive support conducting a job search or identifying financial planning resources. The strategy for recruiting students and others at the Evans Community School remained consistent over the course of the cooperative agreement. Children's Home Society considered students and family members to be enrolled when they received services at the Evans Health and Wellness Center.

The awardee modified their strategy for recruiting Pine Hills community residents several times over the course of the cooperative agreement. To identify and recruit individuals from the community during the first program year, the awardee reached out to patients whom WellCare, a local Medicaid MCO, identified as having gaps in their primary care. Health navigators based out of the Pine Hills Wellness Project office then contacted those WellCare patients and connected them with services as needed, including referrals to what was known at the time as the Wellness Cottage. The awardee enrolled community members when they first received services at the Wellness Cottage. In the second year of the cooperative agreement, however, the relationship between Children's Health Society and WellCare weakened and WellCare no longer shared lists of potential participants with the awardee. As a result, health navigators focused recruitment and enrollment efforts on the broader community; this activity continued through the third year of the cooperative agreement.

Although not officially enrolling participants through these activities, the community-based health navigators effectively reached additional members of the community. According to awardee reports, in the 12th program quarter alone, navigators based at the Pine Hills Wellness Project office completed 40 community outreach activities (for example, hosting and attending health fairs, manning tables at local businesses, conducting events at local consulate offices, and initiating promotions on local radio) and completed a total of 166 such activities during the third year of the cooperative agreement. Program staff believed that these community contacts helped to drive enrollment by encouraging individuals to seek appropriate health care at the Evans Health and Wellness Center, where they were then enrolled in the program.

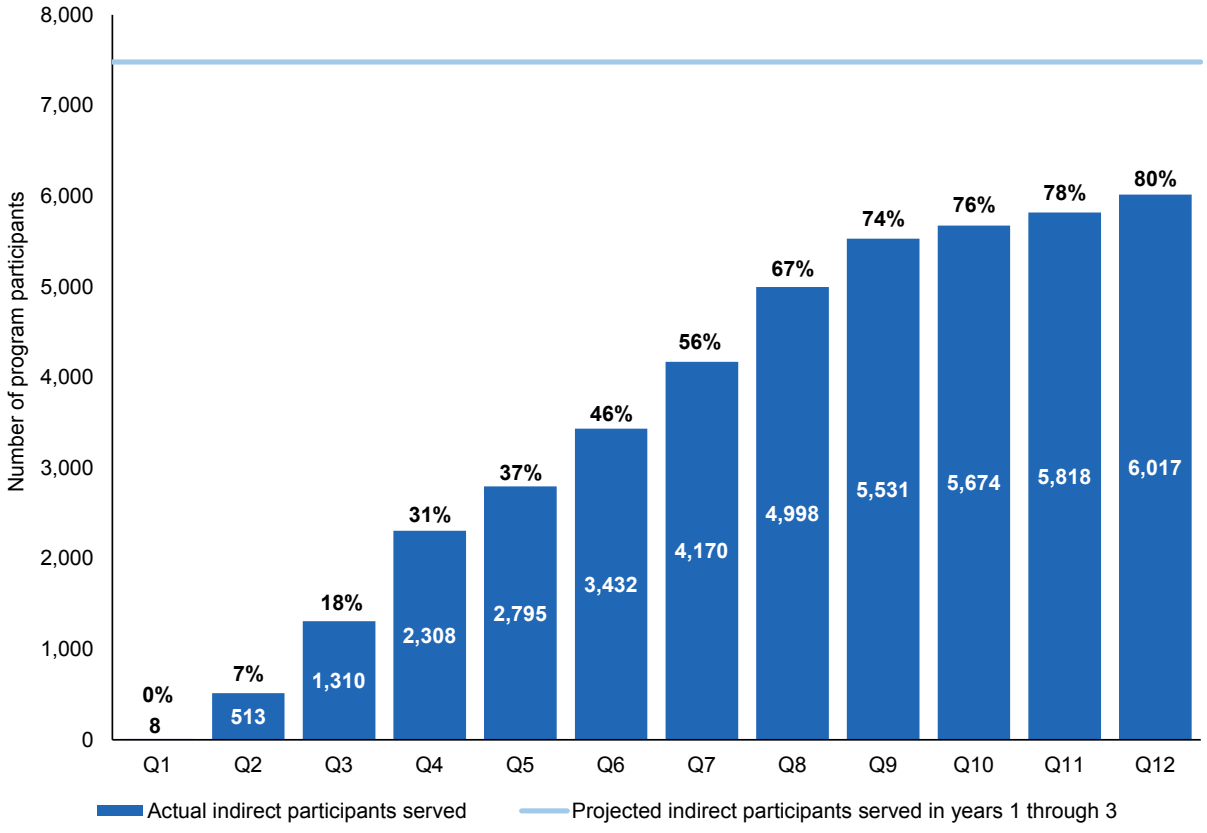
Children's Home Society enrolled students, family members, and community members into the program using the same process. The awardee considered an individual to be enrolled when he or she received patient navigation or medical, dental, or behavioral health care services at the Evans Health and Wellness Center location, and (as explained above) defined all participants as indirect participants. The awardee did not change its enrollment strategy during the three-year cooperative agreement. It is important to note that the awardee temporarily suspended all enrollment activities in the third year of the cooperative agreement during the transition between the True Health and Orange Blossom clinical partners.

b. Evidence of enrollment effectiveness

Overall, the awardee enrolled 6,017 indirect participants from October 2014 (when it launched its program) through August 2017; this represents about 80 percent of its final three-year projections (Figure II.1). Children's Home Society revised its enrollment target several

times during the cooperative agreement, reflecting initial success and subsequent challenges. Specifically, the awardee first set an enrollment target of 1,500, but after surpassing that target in the first program year (enrolling 2,308 individuals), it increased its target to 8,688. After this adjustment, however, enrollment plateaued, and at the end of the second year, it had enrolled 4,998 participants, or 57.5 percent of the new target. The awardee then adjusted its final enrollment target downward to 7,481 in November 2016 to reflect a decrease in service availability (and thus enrollment opportunity) during the transition between its FQHC clinical partners (True Health and Orange Blossom). True Health began ramping down enrollment of new participants starting in August 2016 and halted enrollment entirely for the months of January and February 2017. Once Orange Blossom began operating at the Evans Health and Wellness Center with reduced hours, enrollment did not return to the levels achieved by the awardee prior to the clinical partner transition. Because Children's Home Society did not track enrollment separately for students, family members, and community residents, or the characteristics of enrollees, we do not know if the awardee enrolled the types of participants it originally intended.

Figure II.1. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers. Children's Home Society revised its

Figure II.1 *(continued)*

enrollment target several times during the cooperative agreement, reflecting initial success and subsequent challenges. Specifically, the awardee first set an enrollment target of 1,500, increased its target to 8,688 after Year 1, and later adjusted its final enrollment target downward to 7,481 in November 2016.

c. Barriers and facilitators associated with enrollment effectiveness

The Children's Home Society's progress in meeting its three-year enrollment goals was facilitated by activities during the first and second years of the cooperative agreement that allowed the awardee to overcome initial challenges related to operating on a school campus, ultimately turning its partnership with the Orange County Public Schools (OCPS) and its location at the Evans Community School into assets. Nonetheless, Children's Home Society experienced additional barriers to enrollment related to navigation staff turnover and the transition between clinical partners.

Specifically, the location of the program on the grounds of the Evans Community School fostered access to and thus enrollment of students, in spite of initial conflicts between the partners' priorities that complicated enrollment early in the cooperative agreement. Generally, cooperative efforts between program staff and OCPS staff allowed for intensive, tailored outreach to students. In addition, the Student Ambassadors program encouraged enrollment through peer-to-peer outreach among students and remained popular among students throughout the cooperative agreement. Interviewees reported that students were typically more trusting of the program than community residents and would often bring information about available services home to their families—thus encouraging enrollment of family members. However, in the first program year, ongoing negotiations between OCPS, True Health, and the program were necessary to balance the competing priorities of student education and program implementation. For example, staff related in interviews that stakeholders had to negotiate the best way to foster student availability for medical appointments without interfering with schoolwork. Over the course of the three-year cooperative agreement, the program developed specific clinic times dedicated for students and focused on increased buy-in and participation from teachers, which together enhanced students' abilities to enroll. By the second program year, the three partners had agreed on school policies that fostered student enrollment and minimized classroom disruptions.

However, the awardee also faced barriers to enrollment effectiveness. Substantial turnover among health navigator staff during the second program year, for example, hindered recruitment and enrollment of Pine Hills community members. This staffing interruption left the community-based health navigation team consistently short-staffed in the second year of the cooperative agreement. Both program staff and navigators reported in interviews that navigators had trouble balancing the workload of door-to-door recruitment activities with navigation support to the growing number of enrolled participants. To address the issue, the awardee hired replacement navigators who had experience in social service navigation through employment at another Children's Home Society program. In addition, community-based health navigators started to use more efficient recruitment methods. For example, rather than door-to-door outreach, Children's Home Society worked to build partnerships with businesses and institutions already trusted by the community, and then hosted recruiting events within the community at those institutions.

The transition to a new clinical partner at the Evans Health and Wellness Center in the third program year also posed a significant barrier to the enrollment of both students and community

members. As relayed in interviews, the original clinical partner, True Health, failed to provide services to (and thus enroll) a sufficient number of patients through the school-based clinic to make the partnership with the awardee financially sustainable. True Health's departure completely halted the provision of medical and dental services as well as enrollment efforts for several months. Although awardee leaders and program staff believed the replacement clinical partner, Orange Blossom, was a better fit for the program given its established relationships within the target population, it was not yet fully operational at the Evans Health and Wellness Center by the end of the cooperative agreement. Thus, enrollment activities were not fully ramped up.

Finally, community members' mistrust of institutions hindered Children's Home Society's progress in meeting its three-year enrollment goals. Specifically, staff and leadership consistently reported in interviews that community members' lack of trust in government and community service institutions hindered enrollment efforts throughout the cooperative agreement, though it was most pervasive in the first program year. In the staff survey, for example, most respondents reported that resistance to the program from participants was a barrier to the program achieving its enrollment goals. To help overcome the community's mistrust, the awardee developed and leveraged relationships with trusted and knowledgeable community organizations as outreach and recruitment partners. For example, the awardee partnered with local businesses and organizations like day care centers and local consulate offices for Mexico, Haiti, and Colombia to conduct outreach activities. In interviews during the third year of the cooperative agreement, program staff described these organizations' and businesses' expertise on community needs and priorities as "critical" to their recruitment and enrollment efforts.

Hiring community members to serve as navigators also partly helped overcome resident mistrust of the program. According to interviewees, hiring community members to serve as navigators supported the awardee's efforts to enroll Pine Hills community residents, because residents were more receptive to invitations to enroll in the program offered by their peers. The awardee also attempted to build trust among community residents by focusing its community outreach events on the program's accessibility and inclusivity. The navigation staff held outreach events after work hours in geographically accessible settings and often provided food and child care to facilitate attendance.

2. Delivery of program services

a. Description of and changes to service delivery model

Overall, the Children's Home Society expected that the patient navigation and medical, dental, and behavioral health care services provided through the HCIA R2 cooperative agreement would lead to lower costs of health care, better use of appropriate health care services, and improved outcomes for participants. The awardee also expected that participant engagement activities through the Student Ambassadors program would contribute to these outcomes.

Key components of the Children's Home Society service delivery model included:

- Offering medical, dental, and social/behavioral care at the Evans Health and Wellness Center to meet students', families', and community members' unmet needs

- Offering behavioral health services to students on the Evans Community School campus at the Hub
- Assigning navigators to family, student, and community participants to help them access needed services and resolve the issues they presented with through the Hub and the Pine Hills Wellness Project
- Using the Student Ambassadors program to provide health education services to student participants and activate participants to share health information and encourage participation in program services among their peers

The Children's Home Society reasoned that participants who accessed program services would receive health education, preventive health care services, and support to help manage chronic illnesses. Health education, preventive services, and chronic illness supports, the awardee hypothesized, would lead to positive lifestyle changes that would improve health outcomes and reduce participants' need for higher cost inpatient and emergency services. The awardee believed that basing program services on the Evans Community School campus would increase the potential impact of the intervention. It expected that students would be receptive to program services and that it could leverage students' receptivity to, and positive experiences with, the program to engage their families' and other community residents' participation in the program.

The awardee did not significantly change its service delivery model during the course of the three-year cooperative agreement. It did, however, change the make-up of its partners in the second program year. At that time, Children's Home Society partnered with a new FQHC, Orange Blossom, after the departure of the original clinical partner, True Health. The Children's Home Society discontinued provision of dental care once the Health and Wellness Center re-opened after the transition; although the awardee hoped to reinstate dental service after Orange Blossom had settled into its new role, there were no plans in place.

b. Evidence of service delivery effectiveness

Based on our analysis, Children's Home Society was partly effective in delivering program services. The awardee's successes included delivering services as intended for most participants (with the exception of the time during and after it changed clinical partners, when the transition interrupted medical and dental services), applying lessons learned to help them successfully bring a new clinical partner onboard, engaging community providers and organizations and engaging students in health care services and the Student Ambassadors program, and slowly building relationships with the community. The awardee struggled with connecting with and maintaining trust among community residents, navigation staff turnover in the early years of the cooperative agreement, and a lack of transparency and communication issues with its initial clinical partner. We provide details in the sections below.

Delivery of intervention services. The Children's Home Society was partly successful in delivering navigation, medical, dental, and behavioral health services as intended. Interviewees reported that navigation staff at the Hub and Pine Hills Wellness Project office were largely able to help participants access needed services, particularly among students at the Evans Community School. Respondents to the staff survey also indicated that navigators educated participants about

how to manage their health care and access services, and assisted participants with accessing nonmedical services (for example, housing, job training, and benefit programs). Health navigators provided these services starting in the first program year and did so throughout the cooperative agreement. The Children's Home Society also largely succeeded in delivering behavioral health services to students at the Hub through a counselor available on-site and by appointment.

The awardee struggled, however, to consistently offer medical and dental services at the Evans Health and Wellness Center, most notably due to the transition in its FQHC clinical partner during the second program year. Because of this transition, the FQHC temporarily halted all medical services, and discontinued dental services indefinitely during the third program year. The awardee also reduced the Evans Health and Wellness Center's hours of operation during the third program year, further limiting the awardee's ability to provide services as originally intended. Interview respondents did not have a firm timeline for reinstating the full set of medical and dental services or for offering expanded hours.

In addition, the awardee provided health education on topics such as nutrition, dental care, self-care, and exercise to students at the Evans Community School through the Student Ambassadors program. Interviewees said that students who participated in the Student Ambassador program engaged their peers through informal surveys of current health behaviors then compiled and discussed results of surveys with the larger student population. Additional activities included field trips to local hospitals, gyms, and other health-related places. A member of the program leadership believed that the Student Ambassadors component of the program had strengthened throughout the course of the cooperative agreement, observing that there were more students engaged in the third program year than in prior program years. Interview respondents estimated that somewhere between 15 and 30 students participated in the program during the third year of the cooperative agreement.

Staffing and training. The Children's Home Society was partly successful at meeting its staffing goals. Across interviews conducted during the third program year, staff consistently reported that the program was adequately staffed and had not suffered service delivery challenges related to staff turnover. Program staff included health navigators at the Pine Hills Wellness Project office and the Hub on the Evans Community School campus, a behavioral health provider at the Hub, and leadership and administrative staff. Evans Health and Wellness Staff were employees of the FQHC clinical partner and therefore not considered to be program staff.

Although the program experienced turnover among its navigation staff that hindered enrollment efforts (as described previously), the awardee filled navigator vacancies quickly to minimize service delivery disruption. In addition, interview respondents noted that navigation staff turnover had the unanticipated benefit of replacing original staff with individuals who were better suited to working with the target population, given their membership in the community and their work experience engaging young people. Though the awardee also experienced turnover in project leadership in the third year of the cooperative agreement, the new manager was promoted from within Children's Home Society; interviewees reported that this staffing change did not affect program implementation.

In addition, most staff who completed the survey reported feeling adequately trained for their work and said that they had the time and resources necessary to deliver program services. Staff also indicated they felt supported in their work by their colleagues and management. Nearly all respondents reported satisfaction with their role in the program.

Engagement of providers. Although the Children's Home Society was successful in engaging external community providers and organizations, it struggled to attain the level of engagement that it desired with its clinical delivery partners. Interviewees reported that the program had developed strong relationships with many of the community-based organizations to which it referred participants for needed services. Program leadership interviewees and partners noted, however, that the awardee faced issues with transparency and communication with True Health from the beginning of the partnership, including True Health's reluctance to participate in data collection. True Health eventually dissolved its relationships with the awardee in July 2016. In interviews during the third program year, school-based staff provided mixed reports of their ability to adequately engage the replacement clinical partner, Orange Blossom. For example, the awardee was unable to engage Orange Blossom in activities outside the clinic itself, citing in particular its absence from after-school programs such as the Student Ambassadors program.

Engagement of program participants. The Children's Home Society was mostly successful in engaging student participants, but struggled to engage community participants. Although all staff who completed the survey in the third program agreed or strongly agreed that the program successfully engaged participants in the services provided, staff who participated in interviews consistently reported challenges to engaging community residents, many of whom mistrusted government and community institutions, they said. Staff who completed the survey also reported that resistance to the program by participants was a barrier to the program achieving its goals. Further, when asked why potential participants would opt out of participation, the most common responses were that the participant did not want to make a change (because the participant was happy with the existing system of care or for other reasons) or had concerns over the financial burden or insurance coverage. In interviews, school-based staff consistently reported that students were generally more engaged with the program than community residents, especially through the Student Ambassadors program.

c. Barriers and facilitators associated with service delivery effectiveness

Children's Home Society's ability to effectively deliver program services on the Evans Community School campus was facilitated by (1) Pine Hills Wellness staff's ability to offer programs that excited students and (2) the commitment of awardee leadership.

First, Pine Hills Wellness was able to engage students in program activities by incorporating elements of health and wellness into games (such as scavenger hunts) and giveaways (such as snacks). Interview respondents noted that because student participation was voluntary, these sorts of compelling activities fostered student participation.

Second, respondents identified Children's Home Society leadership's commitment to improving the lives of students at the Evans Community School as a facilitator of program implementation. This commitment, they said, was evident to the students and the school administrators, and respondents observed that it positively influenced the campus climate and helped to build a solid partnership between the Children's Home Society and OCPS.

The awardee also faced several barriers to service delivery effectiveness, some of which it overcame more successfully than others. Barriers to service delivery effectiveness included (1) transition in clinical partners, (2) community mistrust, (3) the nature of service delivery on a school campus, (4) the limited resources of community participants, and (5) safety issues in the surrounding neighborhood.

First, though the Children's Home Society's ability to delivery services on the Evans Community School campus and to the Pine Hills community was clearly hindered by clinical partner turnover, the awardee was able to overcome some of the challenges faced by its initial clinical partner, True Health, by applying lessons learned when bringing Orange Blossom on board as its replacement partner. Upon receiving notice that True Health planned to terminate its partnership in July 2016, awardee and program leadership conducted an examination of their successes and shortcomings from the prior year. During the search for a new FQHC clinical provider with which to partner, leaders deliberately considered the lessons learned to determine desirable attributes and the best way to proceed with prospective partners to ensure a more durable partnership. Interview respondents noted, for example, that the awardee engaged with Orange Blossom earlier in the process than they had with True Health. Efforts to bolster the partnership included early meetings about the logistics of working with students, as well as a revised approach to implementation that focused on starting small and building trust, rather than immediately offering more services and operating during more hours than was sustainable (as was experienced by True Health, which struggled to adapt its service provision at the Evens Wellness Cottage to support the needs of student and community populations in a financially viable way). Respondents also expressed optimism, believing Orange Blossom was well suited for partnering with the Children's Home Society and the Pine Hills Wellness Project because of its background in providing care for underserved populations.

Second, staff reported ongoing challenges related to community residents' lack of trust and willingness to engage with the program. This challenge was first described by staff in interviews during the first year and was repeated each year thereafter. According to the awardee, because the community is predominantly made up of low-income households and includes a large immigrant population, many organizations receive grants to implement short-term programs, and then close their doors when the funding ends. This cycle has fortified residents' mistrust of government-funded programs such as the awardee's. The awardee worked to overcome this challenge by providing navigation services and meeting community members where they were, at community events, local businesses, and anchor institutions such as churches, and to establish itself as a community-based organization that was there to support and coordinate services for those in need. The awardee also established and leveraged a community advisory board to reach out to community providers, build community trust, and expand its referral network. In the second year and third years of the cooperative agreement, program staff described slow progress in gaining the trust of the community, describing how the awardee's reputation was spreading within the community, and community-based organizations such as Goodwill Industries International, Inc. had reached out to program leaders to develop partnerships.

Third, the location of the program on a school campus also affected Children Home Society's ability to effectively deliver services, much the way it affected enrollment. Specifically, there are inherent limitations to staging a health intervention at a school, including challenges scheduling programming when school is not in session, difficulty balancing the

benefits of health services and the drawback of pulling students from their classrooms, and logistical barriers to access for community members who are not affiliated with the school (for example, Florida state law prohibits student and community populations from accessing the school-based clinic during the same times). Yet, colocating services on the campus in many ways facilitated students' access; staff interviewed during the third program year largely discussed the benefits of the arrangements they had negotiated with OCPS and Evans Community Schools during previous years that now fostered student access to services while also supporting student education.

Fourth, respondents noted that service delivery to community members was hindered because some residents lacked the resources necessary (transportation availability, funds for bus fares, time given work requirements) to attend preventive or follow-up health appointments at the Evans Health and Wellness Center. For example, respondents suggested that despite their efforts to educate community participants about the importance of routine health maintenance visits and to inform community participants of available bus routes to help them access care at the Evans Health and Wellness Center, some participants would likely continue to received services at an emergency department, as opposed to the Center, due to its relative convenience.

Finally, the accessibility and safety of the Health and Wellness Center was a concern that impeded service utilization through the entire cooperative agreement. Though administrators assumed that the convenient location of the Evans Wellness Cottage would promote service use, they found that the Pine Hills community faces several barriers to accessing the Cottage. For instance, there is no direct route from the bus depot to the Cottage, so patients must walk through a dangerous neighborhood to access services. The Children's Home Society continued exploring options for addressing this challenge, such as building an access road. At this point, however, accessing the Evans Wellness Cottage continues to be a challenge. Until True Health's departure from the program, they had worked with the awardee to improve signage and provide transportation information to prospective patients. It was not clear how these efforts would change, if at all, with Orange Blossom's presence.

C. Assessment of perceived program effects on the delivery of care and outcomes

> "Looking at the efforts that [Children's Home Society has] put in across the board, I would think that there should be some evidence [that community health has improved]."
>
> —Awardee partner

In interviews, awardee leaders and staff were largely reluctant to comment on the likely effects of the program on the delivery of care or outcomes. Although most leaders and staff perceived that that both student and community participants appreciated and enjoyed program services, they provided only scant anecdotal evidence suggesting that community health improved as a result. Several frontline and administrative staff attributed the potential success of the intervention to the educational aspects of the health navigation and the Student Ambassador program, including the dissemination of information on how to obtain direct care services.

Respondents to the staff survey, however, were more forthcoming in providing examples of perceived impacts of the program. For example, all staff responding to the survey reported that the program had been somewhat or very effective at achieving its goals, and nearly all agreed

that the program was making a difference in meeting critical needs in the community. Nearly all staff indicated that the program had a positive impact on the achievement of the participants' health goals, with one respondent describing the program as "contributing to the betterment of the community," with its effects "seen in tangible results."

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Three implementation factors will likely challenge the design and interpretation of an analysis of the Children's Health Society's impact on desired outcomes.

First, in light of the awardee's inability to meet its enrollment target, it is unlikely it was able to affect community-wide outcomes such as the number of emergency department visits per beneficiary. In addition, because the awardee did not differentiate its enrollment numbers for different populations (such as students versus community residents), it is unclear if school-based outcomes, such as the percentage of female participants younger than 18 who are pregnant or the percentage of youth enrolled at Evans Community School who report risky health behaviors, are more or less likely to be affected.

Second, there was difficulty in providing medical, behavioral, and dental services over the course of the cooperative agreement, due to initial logistical and scheduling difficulties as well as the transition between clinical partners in the second year of the cooperative agreement. These issues limited participants' exposure to critical components of the intervention and likely mitigated its impact on clinical outcomes.

Finally, the intervention relies on several funding streams, making it difficult to isolate the effects of HCIA R2 in any impact evaluation.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Children Home Society's program.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data. Children's Home Society of Florida's program ended August 31, 2017.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: Children's Home Society of Florida

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure by February 28, 2017	0[a]
Projected Medicaid population with 6 months of program exposure by February 28, 2017	Unable to project at this time due to lack of Medicaid data
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	4,582
Likelihood of all-cause hospitalizations	2,932
MDE sample size requirement to detect 20% effect	
Total expenditures	1,146
Likelihood of all-cause hospitalizations	733
Participation/Selection bias of concern	Yes, patient self-selection high or high refusal rate
Intervention implemented that differs from baseline period	Questionable, patients may have been receiving intervention prior to HCIA R2 cooperative agreement
Claims sufficient to identify treatment and comparable comparison group?	Questionable, no testing yet to determine strength of intent-to-treat framework
Likelihood of solid comparison group	Too early to determine due to delay in Medicaid data
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Lack of timely Medicaid data by final report
Survey data for treatment group that will be analyzed	Staff survey
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

At this point, it is unlikely that we will be able to conduct a rigorous impact evaluation of the awardee's program for two reasons: the Alpha-MAX Medicaid data for Florida are lacking Medicaid managed care data, and we have received only MCOs' patient identifiers, which do not allow us to link to Alpha-Max or the future T-MSIS Medicaid data. We continue to explore the feasibility of obtaining Medicaid enrollment and encounter data from the one Medicaid MCO that has the largest share of the awardee's participants. We have signed a business associate agreement with the MCO and are now waiting for enrollment data for all of its enrollees who have participated in the awardee's program and for a potential comparison group. Upon receipt of the enrollment data, we will determine whether we are able to link the identifiers provided by the awardee to the enrollment data, calculate the number of enrollees with usable data during the baseline and the intervention period, conduct a power assessment, and update our assessment of evaluability. At that point, we will determine whether to obtain encounter data for the treatment and comparison groups. If managed care encounter data do not become available, we will not be conducting an impact analysis and will report only on the experiences of awardee staff and participants, based on our surveys.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Children's Home Society's payment model was still in development. Children's Home Society envisioned a per beneficiary per month (PBPM) fee from Medicaid MCOs that would cover primary and preventive medical, dental, outpatient behavioral health care, and community health and wellness promotion activities.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: (1) the awardee's final payment model report, submitted in June 2017 to the implementation and monitoring contractor and (2) narratives and payment model reports the awardee submitted to the implementation and monitoring contractor. We did not interview the awardee about its payment model because it did not advance from the development stage to active negotiations with one or more payers.

B. Description of the payment model

Under the proposed payment model, Medicaid and Children's Health Insurance Program (CHIP) MCOs in Florida would pay a PBPM fee to Orange Blossom, which would then subcontract the provision of the covered primary and preventive care services to Children's Home Society.

The population covered by the payment model includes all residents of Pine Hills, Florida, (zip codes 32808 and 32818) enrolled in Medicaid or CHIP. The services covered by the payment model include primary and preventive medical, dental, outpatient behavioral health care, and community health and wellness promotion activities.

The payment amount would be negotiated with an MCO. The awardee reported that payment amount would be based upon a number of factors, including the cost of providing the services, cost savings as measured by changes in emergency department and inpatient costs, quality of care as measured by CAHPS surveys, and quality measures (using HEDIS). However, the awardee had not reached the point of specifying these adjustments on its own or in negotiations with MCOs.

C. Status of the payment model

The awardee had not begun any negotiations with payers. However, the awardee planned to continue talking with local hospitals in order to explore collaboration on future payment models.

D. Factors associated with the development of the payment model

The largest barrier the awardee reported to development of its payment model was difficulty gaining access to claims data to assess the impact of the program on utilization, costs, and quality. As of August 2017, the awardee had not been able to reach agreements with MCOs for data sharing. The University of Florida is conducting a qualitative analysis of the awardee that is expected to be completed in December 2017; Children's Home Society hopes it will convince payers of its value.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Despite experiencing challenges finding funding for its program after the cooperative agreement ended, the Children's Home Society of Florida reported that it will sustain some aspects of the program, including a similar program that was newly implemented in the third program year in a nearby neighborhood. The awardee reported that challenges obtaining data prevented it from being able to prove the program's value to engage payers.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

To sustain the program after the end of the cooperative agreement, the Children's Home Society of Florida started exploring funding options in the second program year. Possible funding options included applying for grants, engaging philanthropists, and leveraging relationships with partnering organizations to continue providing program services.

C. Implementing the SSR plan: progress and changes

Sustainability. The awardee reported no definite plans to sustain the entire program after the end of the cooperative agreement, although program leaders expressed belief that some

> "Every community school is a unique space. Every neighborhood, every city [has] different demographics, needs, and resources. [Our program services are tailored to] what [the community] perceives as a need."
>
> —Program leader

elements of the program may be sustained. According to multiple interviewees, the awardee is seeking to continue the Student Ambassadors program, navigation and behavioral health services at the Hub, and medical services at the Health and Wellness Center, but did not have concrete plans for doing so. One respondent noted that in the third program year, all program partners signed an informal agreement that committed partners to continue their partnership for at least 25 years. Because their services are not funded by HCIA R2, Orange Blossom will continue providing health services for the program. "We're committed, we're passionate, and we're dedicated. Someone will always be at the table to make sure we're successful as a partnership," one interviewee said about the lasting potential of program partnerships.

Due to budget constraints, the community-based navigation services offered through the Pine Hills Wellness Project office will not be sustained after the cooperative agreement, according to the awardee. Programs leaders mentioned the possibility of reopening these services through Pine Hills Wellness Project at a later date when funding is available, but noted that eliminating then restarting the program will do irreparable damage to the program's reputation and ability to engage the community, especially in the context of a community like Pine Hills, with established mistrust of institutions reinforced by temporary, grant-funded interventions. "You have a stigma that you're not going to be there when community members need you," a program leader explained.

Scalability. The Children's Home Society expanded its health navigation program in August 2017 to a new site in a nearby, low-income neighborhood similar to Pine Hills. Much like the program's combined navigation services through the Hub and medical services through the Evans Health and Wellness Center on the Evan Community School campus, the new facility implemented a health program that employs a clinical provider and leverages other community health care resources and partners. Other services at the new site may differ in order to meet the needs of the specific community, however.

Replicability. The Children's Home Society did not report plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

The inability to secure funding was the primary challenge to program sustainability, according to awardee leaders. Inadequate data from HMOs and insurance companies prevented the awardee from calculating cost savings attributable to the program; as awardee leaders stated, without such data the awardee could not engage payers or attract funding. The awardee expressed hope that expanding to the new site in August 2017 would allow it to collect more data and build a stronger case for cost savings. The Children's Home Society is looking for grants to sustain the Student Ambassadors program and continue its "positive momentum," but may also be able to operate it as an unfunded after-school club sponsored by a teacher. Finally, the lack of funding made it impossible to sustain many services and positions, according to an interviewee from a partnering organization. In contrast, one-off program activities like health fairs are easier to sustain because they are less costly, the interviewee said.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. However, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or the sustainability, scalability, or replicability plans. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Clifford W. Beers Guidance Clinic, Inc.

April 18, 2018

Michaella Morzuch (Mathematica Policy Research)
Grace Anglin (Mathematica Policy Research)
Lisa Lines (RTI International)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Clifford W. Beers Guidance Clinic Inc., a community-based mental health clinic in Connecticut, received HCIA R2 funding to implement the new program, Wraparound New

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Clifford W. Beers Guidance Clinic, Inc., received a six-month, no-cost extension through February 28, 2018. During the extension, Wraparound New Haven intended to conduct project phase-out and complete analyses of the data on cost savings.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

Haven (Table I.1). The awardee was motivated by the interconnected nature of the behavioral and physical health needs of Medicaid-enrolled children, which were often exacerbated by family stressors such as housing or food insecurity. Wraparound New Haven had two innovative features: (1) an emphasis on addressing both physical and behavioral health needs and (2) a focus on managing and coordinating services for all interested family members. The program focused on recruiting children younger than 18 years old in Greater New Haven who were enrolled in Medicaid, had a chronic physical health condition, had or were at risk of a mental health condition, and had more than one ED visit or any hospitalization in the prior 12 months. Children meeting these criteria and all interested members of their families were enrolled in Wraparound New Haven for 6 to 12 months, during which time care coordinators helped families identify health and wellness goals, prioritize strategies for addressing those goals, and develop a care plan. If needed, Wraparound New Haven's behavioral health clinicians provided bridge counseling services to family members of the primary enrollee until a more permanent source of care was secured.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Wraparound New Haven connected eligible high-need children and their families to care coordinators to improve the management, coordination, and integration of behavioral and physical health services and social supports.
Major innovation	Wraparound New Haven was innovative in that it focused on the entire family unit's behavioral and physical health needs, rather than a single participant's behavioral health needs.
Program components	• Care management services • Integrated behavioral and physical health care services • Participant and family engagement
Target population	Wraparound New Haven provided services to a primary enrollee—children who met the eligibility criteria—and to all members of their families who wished to participate. The primary enrollee had to meet the following criteria:[a] • Resident of Greater New Haven, CT • No older than 17 • Current Medicaid beneficiary • At least one chronic physical health diagnosis (broadly defined as any condition that consistently impacted a child's health status) and one mental health diagnosis or living with conditions that tended to predict mental health conditions • Either two or more ED visits or any hospitalization during the prior 12 months
Theory of change/ theory of action	Clifford Beers Guidance Clinic hypothesized that families working closely with a care coordinator would have a better understanding of how to manage their own health, determine the services they needed, and find those services. Accessing these services would, in turn, result in improved mental and physical health outcomes and lower health care spending.
Payment model	Clifford Beers Guidance Clinic reached a fee-for-service reimbursement arrangement with a major private insurer for Wraparound New Haven services. It was also pursuing a value-based, capitated payment arrangement with the Connecticut State Department of Social Services.
Award amount	$9,739,427
Effective launch date	12/2/2014

Table I.1 *(continued)*

Program characteristic	Description
Program setting	• Community-based • Home-based
Market area	Urban
Market location	Greater New Haven, CT
Target outcomes	• Improve the coordination of physical and behavioral health care • Enhance family engagement with providers • Improve participants' physical and mental health status • Increase participants' social connections and social supports • Reduce health care spending

[a]The awardee, with CMS approval, changed the eligibility criteria in program Year 1 from requiring the primary enrollee to have three or more visits to the ED to two or more visits and expanded the service area from New Haven proper to Greater New Haven (17 cities and towns).

ED = emergency department.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we concluded that Clifford Beers Guidance Clinic was partly successful in implementing its program by the end of the initial three-year cooperative agreement. Clifford Beers Guidance Clinic staff kept participants actively engaged with the program and generally delivered intervention services as intended by the end of program Year 2. However, the awardee's implementation effectiveness was limited by challenges in meeting its enrollment target and community providers' lack of engagement with the program.

Impact evaluation. We plan to conduct a rigorous assessment of the impacts of Clifford Beers Guidance Clinic's program; however, the analysis is still in progress and is not included in this report.

Payment model. Clifford Beers Guidance Clinic reached an agreement for fee-for-service (FFS) reimbursement for program services from a large commercial insurer by the end of the cooperative agreement, but was not able to reach an agreement with the Connecticut Department of Social Services (DSS).

Sustainability plans. The core components of the Wraparound New Haven model will not be sustained for the original focus population of children enrolled in Medicaid. However, the model will be scaled by using the FFS reimbursement contract with a commercial payer in Connecticut. Clifford Beers Guidance Clinic disseminated information about the program to other insurers, but there were no indications that these insurers intend to support the program as originally developed.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey, which was fielded from July 2016 to October 2016 with all non-physician staff members at the time (33 respondents), achieved a response rate of 90 percent. The clinician survey, which was fielded from March 2017 to June 2017 with the four physicians on staff at that time, achieved a response rate of 100 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items with fewer than 11 respondents, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we concluded that Clifford Beers Guidance Clinic was partly successful in implementing its program. Although the awardee increased enrollment over time, it fell short of its enrollment goal of 2,284 participants, instead enrolling 1,944 participants (primary enrollees and their family members) by August 2017. Program staff reported that participants were actively engaged with the program. Although care coordinators initially faced challenges in addressing participants' physical health needs, the awardee ultimately achieved its goal of addressing families' physical and behavioral health needs. However, the awardee was less successful in engaging community providers, other organizations, and other individuals who supported participants, such as physical and mental health care providers, social service and community organizations, and family friends and neighbors. Clifford Beers Guidance Clinic received a six-month, no-cost extension, which was scheduled to end in February 2018.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

To identify potential participants, Clifford Beers Guidance Clinic initially relied solely on faxed or mailed referrals from community health care providers and social service organizations. Despite the awardee meeting with referral providers to explain the program, referral rates remained low. Thus, at the end of program Year 1, the awardee made the following two changes to its participant identification and referral process:

1. The awardee expanded the program's eligibility criteria, viewed by program staff and referral providers as initially too restrictive. Specifically, the awardee relaxed the prior utilization criteria for the primary enrollee from three to two emergency department visits and expanded the residency criterion from New Haven proper to Greater New Haven (17 cities and towns).
2. Awardee staff started directly identifying potential enrollees in addition to receiving referrals. In October 2015, the awardee embedded a nurse at a major potential referral site, the Yale New Haven Hospital Primary Care Center (PCC). The nurse reviewed the electronic medical records of children with appointments at the hospital who had an ED visit or inpatient admission in the prior 12 months, to identify those potentially eligible for Wraparound New Haven. The nurse then met with families before or after their child's appointment to describe the program, gauge interest, and determine whether a given family was a good fit for either Wraparound New Haven or a less-intensive care coordination program offered by Clifford Beers Guidance Clinic and other community providers. The nurse also educated Yale New Haven Hospital clinicians about the program to encourage referrals.

After community organizations or the embedded nurse referred families to the program, the Wraparound New Haven intake coordinator verified potential participants' eligibility. If deemed eligible, a lead care coordinator assigned the potential participant to an available care coordinator who spoke the family's primary language (when possible) or who had specialized expertise that matched the family's perceived needs. This care coordinator was responsible for following up with the family and scheduling a first appointment to introduce the program. Depending upon family preferences and availability, first appointments could take place by phone or in person at the participants' home, a convenient community location, or Clifford Beers Guidance Clinic. If the family was interested in participating, care coordinators then enrolled the child identified as the primary enrollee and all interested family members.

Families were enrolled in the program for between 6 and 12 months. The primary enrollee remained enrolled for this entire period; their family members could enroll or disenroll at any time.

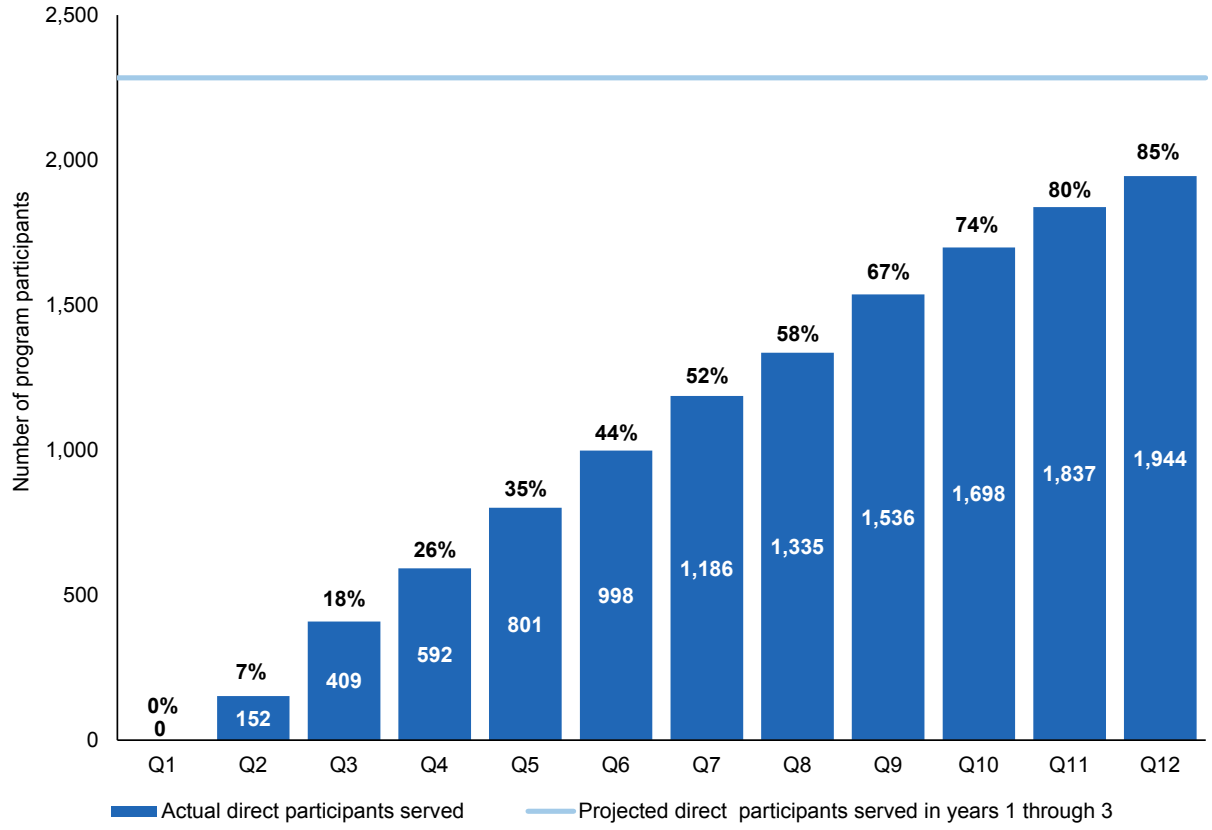
b. Evidence of enrollment effectiveness

Clifford Beers Guidance Clinic was partially effective in reaching its three-year enrollment goals. Over the course of the cooperative agreement, Wraparound New Haven staff reported that the program enrolled children who met the level of physical and mental health needs that the awardee hoped to address. According to self-reported information submitted to the implementation and monitoring contractor, primary enrollees had visited the ED a mean of 3 times in the prior 12 months and had a range of chronic conditions, including asthma (45 percent), depression (21 percent), and obesity (14 percent).[4] However, the awardee fell short of reaching its enrollment goal. Clifford Beers Guidance Clinic reported that it enrolled 1,944 direct participants from December 2014 (when it launched its program) through August 2017, which represents about 85 percent of its final three-year projection of 2,284 direct participants (Figure

[4] Data were self-reported by the awardee on 515 primary enrollees in May 2017.

II.1). (During the cooperative agreement, Clifford Beers Guidance Clinic increased its three-year enrollment target by 1.5 percent—from 2,250 participants to 2,284 participants.) Enrollment was lower than expected during program Year 1 due to difficulty generating referrals from community providers but increased in program Years 2 and 3.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. During the cooperative agreement, Clifford Beers Guidance Clinic increased its three-year enrollment target by 1.5 percent—from 2,250 participants to 2,284 participants.

c. Barriers and facilitators associated with enrollment effectiveness

Two major barriers impeded enrollment in Wraparound New Haven: (1) restrictive eligibility criteria for the primary enrollee and (2) challenges with engaging referral providers. Though less significant, lack of family interest in the program also posed a barrier to enrollment. Clifford Beers Guidance Clinic was able to mitigate these challenges by embedding a nurse in Yale New Haven Hospital to help identify participants and generate referrals.

In the first program year, Clifford Beers Guidance Clinic and its referral providers were concerned that they would not be able to identify enough eligible children to reach their recruitment goals. Wraparound New Haven staff indicated that expanding the eligibility requirements—to children with two ED visits in the prior 12 months as well as children in the Greater New Haven area—expanded the pool of potentially eligible children. However, staff noted that the eligibility criteria remained fairly restrictive and that the program may have reached its enrollment goal by further expanding eligibility to children who had either one behavioral health condition or one physical health condition and one ED visit during the prior 12 months.

Wraparound New Haven received fewer referrals from community providers and social service agencies than anticipated. The awardee reported that competing demands on providers' time limited their engagement with the program. Moreover, the awardee indicated that some providers stopped making referrals to the program after experiencing frustrations with the enrollment process. For example, program leaders said some providers became frustrated after they spent time explaining the program to a family and making the referral only to learn later that the family did not meet the eligibility criteria or had been referred to the program in the past and declined participation or disenrolled. In addition, program leaders reported that one community provider organization may have stopped referring children because the organization was expanding its own behavioral health service offerings, and may have viewed Clifford Beers Guidance Clinic as a potential competitor.

> "Referrals may be sent, and when the patients do not qualify, then the providers get frustrated. Even if he or she spent an extra five minutes talking about Wraparound New Haven to a family, it's five more minutes, and then they don't qualify. And after the provider does that three or four times, he or she gets frustrated, which is why I think a lot of the outside referrals have not worked."
>
> —Program staff

Though viewed as a less significant challenge, awardee staff also reported that some eligible families decided not to meet with a care coordinator following their referral to the program. On the 2016 non-clinician staff survey, respondents were asked if various factors contributed to families deciding not to join the program. Two-thirds of staff reported that families declined participation because Wraparound New Haven required too much of a time commitment (67 percent) or had too many requirements, such as attending care team meetings (63 percent). In addition, 41 percent of staff reported concerns about privacy or confidentiality or had a lack of understanding of the program.

> "With this population, there are a number of unknowns as to why a family continues or not. They may sign the paperwork but do not want an outsider coming into their home. They dislike the care coordinator. Or, they may say yes because they are uncomfortable saying no."
>
> —Program staff

The nurse embedded at Yale New Haven at the end of program Year 1 helped Wraparound New Haven mitigate these challenges and increase enrollment in program Years 2 and 3. Specifically, program staff reported that the nurse's in-depth knowledge of the program's eligibility criteria, direct access to the hospital's electronic medical records, and understanding of physical health providers' workflows all facilitated her ability to identify potentially eligible children and orchestrate "warm handoffs" to Yale New Haven providers and Wraparound New Haven care coordinators. Although time-consuming, the nurse also met with potentially eligible families to fully understand whether a child was eligible and a good fit for the program. For example, the nurse met with the family of a child who, based on the chart review, met all eligibility criteria other than the mental health criteria. During the meeting, the nurse learned that the child also had a mental health diagnosis.

> "I usually don't disqualify many people until I sit down with them. It's really about talking to the families. I find out there's a lot more going on behind the scenes [than is in the chart], and they would be appropriate for Wraparound. This is a hard way to go about it. But I find in this setting the providers have a limited relationship or background with the family for the most part, because patients are seeing different providers and there's not the same relationship as with one pediatrician regularly seeing a child."
>
> —Program staff

2. Delivery of program services

a. Description of and changes to service delivery model

This section briefly describes how the awardee envisioned Wraparound New Haven implementation and highlights changes to the service delivery model. For the first few weeks following enrollment, care coordinators tried to meet with families twice a week to develop a care plan to address the physical, behavioral, and social service needs for the primary enrolled child and all enrolled family members. To inform the care plan development, behavioral health clinicians attended early meetings with families and completed a series of assessments to identify families' physical, behavioral, and psychosocial risks, needs, and strengths. Clinicians aimed to complete all baseline assessments by the fourth home visit. In addition, Wraparound New Haven's medical staff conducted chart reviews soon after families enrolled to help care coordinators understand participants' physical health needs. With the family's permission, the care coordinator shared care plans with and sought input from providers who were treating the participants.

After the care plan was developed, care coordinators tried to meet with families three times per month in person and two times per month by phone to help them achieve the plan's goals. Care coordinators attempted to empower families in their navigation of the physical and behavioral health systems, not to manage those services for them. Specifically, in ongoing meetings, care coordinators aimed to (1) identify and help families address immediate crises, (2) educate families on the health and social service systems, (3) support families' efforts to manage and coordinate services, (4) identify and help mitigate environmental factors in the home that may impede progress towards goals, and (5) provide referrals to community-based services and supports (such as housing or employment assistance). Every 30 to 45 days, the care coordinators attempted to engage the community providers and other organizations or individuals that supported the participants, including physical and mental health care providers, social service and community organizations, and family friends and neighbors. The purpose of the meetings—

referred to as Child and Family Team meetings—was to gain input for the care plan and support for achieving the family's aims.

In addition to working with the care coordinators, participants also received services as needed from other members of the Wraparound New Haven team:

- Wraparound New Haven's behavioral health clinicians provided short-term counseling to family members and helped them secure long-term counseling from other providers as needed.
- The program nurse conducted follow-up calls with families after a family member had an ED visit or inpatient hospitalization to ensure that a follow-up appointment was scheduled and to identify strategies to help the family avoid another ED visit or hospitalization.

The awardee made several changes to its service delivery model to improve the existing services that were offered to participants:

- During program Year 1, the behavioral health clinicians took over responsibility for completing the family assessments due to concerns that the care coordinators were not consistently administering and scoring those assessments.
- During program Year 2, the nurse embedded at Yale New Haven took over responsibility for conducting follow-up calls with families after a visit to the ED or a hospitalization. Awardee leaders indicated that the nurse was better prepared to conduct those calls than the care coordinators, due to her medical background and access to Yale New Haven's electronic medical records.

> "I think this process is working really well because it gives the family that extra support of a nurse from our program calling the family personally and asking, Are you okay? Do you need anything else? And then we follow up with the family right after that, too."
>
> —Care coordinator

- Because providing the same level of service for all participants regardless of need was burdensome for both staff and participants, the awardee created two physical health tracks during program Year 2—one for participants with complex physical health needs (which included primary enrollees and their family members with complex needs who opted in) and one for those without complex physical health needs. Participants in the comprehensive physical health track were supposed to receive the following additional services: ongoing medical record reviews and case reviews by Wraparound New Haven staff physicians and monthly check-ins between care coordinators and participants' primary care physicians.

In addition, the awardee made several changes to its service delivery model to expand the services that were offered to participants:

- In program Year 1, Wraparound New Haven behavioral health clinicians started providing short-term counseling services after the care coordinators indicated that finding counseling services for participants was challenging.
- In program Year 3, the awardee hired a nutritionist after recognizing that participants with hypertension, diabetes, and high blood pressure needed additional support. The Wraparound

New Haven nutritionist worked with families to address physical health conditions related to obesity by, for example, attending home visits, conducting grocery store walk-throughs, and hosting family cooking classes to encourage healthier eating habits.

b. Evidence of service delivery effectiveness

Delivery of intervention services. By program Year 2, Wraparound New Haven staff were generally delivering intervention services to participating families with the intended frequency and focus. According to self-reported information submitted to the implementation and monitoring contractor, care coordinators met targets for the frequency of meetings with families by meeting with them a median of 3 times per month in person and a median of 2 times per month over the phone. The number of in-person visits increased over the course of the program from a median of 2.4 visits per month in program Year 1 to 3.3 visits per month in program Year 3, demonstrating care coordinators' improved adherence to program guidelines. In addition, the awardee reached 83 percent of participants who qualified for follow-up calls after a family member had an ED visit or inpatient hospitalization between January 2016 and August 2017, but did not reach its goal of 100 percent.

Regarding the focus of intervention services, care coordinators largely worked with families on their behavioral health needs in program Year 1—instead of focusing on both behavioral and physical health needs. However, following additional training and support on physical health, care coordinators and awardee leaders reported that the program was successfully addressing both participants' physical and behavioral health needs by the middle of program Year 2. However, the awardee reported that some care coordinators started actively managing services to meet the physical health needs of the families, rather than empowering them to do so themselves.

Staffing and training. Clifford Beers Guidance Clinic hired, trained, and retained enough staff to implement Wraparound New Haven. Program staff reported that they were able to meet their goal of care coordinators serving a median caseload of 10 families or fewer. On the 2016 non-clinician staff survey, most staff said that they were very satisfied (48 percent) or somewhat satisfied (33 percent) with their role. However, staff also frequently reported that their workload was much too heavy (12 percent) or somewhat heavy (72) and that the program was increasing feelings of burnout (67 percent). Toward the end of the cooperative agreement, staff departures increased, which program staff attributed to concerns about program sustainability following the end of the HCIA R2 funding.

All Wraparound New Haven team members received 100 hours of training on the wraparound model of care and common chronic physical health conditions, such as asthma, diabetes, and obesity, before they began working with families. Most staff surveyed reported positive views of the trainings on the non-clinician staff survey: 88 percent reported that they learned new skills important for their role and 81 percent reported that training helped improve their job performance. During the qualitative interviews, staff reported that the additional training on physical health, which was added in program Year 2, was particularly useful.

Recruitment and engagement of providers. Clifford Beers Guidance Clinic did not successfully engage families' physical and behavioral health providers in the development and implementation of care plans. The awardee's self-reported measures indicated lower-than-anticipated rates of provider contacts per family per month throughout the cooperative agreement

(actual value versus anticipated value). Most notably, care coordinators held a median of 0.3 Child and Family Team meetings per active family per month—fewer than would be anticipated if providers were engaged in these meetings every 30 to 45 days as planned. (The frequency of Child and Family Team meetings was also affected by scheduling challenges with families.)

Engagement of program participants. Clifford Beers Guidance Clinic was successful in engaging families in Wraparound New Haven. As noted above, the program met its goals for contacts with families. Eighty-one percent of non-clinician staff surveyed in 2016 reported that they strongly (59 percent) or somewhat (22 percent) agreed that the program successfully engaged participants. In addition, 67 percent reported that participant engagement was a major factor that helped Wraparound New Haven in achieving its program goals.

c. Facilitators and barriers associated with service delivery effectiveness

Clifford Beers Guidance Clinic faced three main barriers to implementing Wraparound New Haven: (1) challenges integrating physical health within the wraparound care planning model; (2) overtaxed staff; and (3) difficulties identifying and accessing sufficient community resources to support participants, such as housing assistance.

First, Clifford Beers Guidance Clinic is a well-respected behavioral health organization but it had limited experience with physical health services and it struggled to integrate the two at the outset of Wraparound New Haven. In program Year 1, many care coordinators were uncomfortable with addressing families' physical health needs. They ensured that families had a primary care physician and reminded them to attend appointments, but they did not actively help families manage their physical health. Then, Clifford Beers Guidance Clinic sharpened its focus on physical health services, most notably by hiring part-time medical staff to support the care coordinators—including, a pediatrician (hired in program Year 1); an internist, a nurse, and an additional pediatrician (hired in program Year 2); and a nutritionist (hired in program Year 3). Medical staff educated care coordinators on common chronic physical health conditions, encouraged them to broach physical health topics with families, and held regular and ad hoc meetings with care coordinators to discuss families' physical health goals. Care coordinators said they valued these supports and increased their attention to families' physical health needs over time, though program staff reported that some care coordinators remained more comfortable or more adept with the physical health component than others.

> "Every step we move away from child mental health is one more step outside of Clifford Beers Guidance Clinic's natural comfort zone. And to focus on the child's physical health conditions, I think we have done a good job with that…. The care coordinators are very comfortable talking to the staff physicians about physical health conditions and they are very comfortable asking the parents about those conditions."
>
> —Program staff

Care coordinators also faced challenges with engaging physical health providers, which contributed to difficulties in integrating physical and behavioral health services. Surveyed staff attributed low levels of engagement to providers' apprehension about program requirements, such as additional meetings or documentation (44 percent) and concerns over the time commitment required (41 percent). In program Year 2, Wraparound New Haven's care coordination platform (Essette) was programmed to remind care coordinators to share family

care plans and assessment results with providers in order to encourage provider engagement. In addition, care coordinators reported other strategies that they used to encourage provider participation, including accommodating providers' busy schedules by having them participate in meetings by telephone or emailing summaries of meetings to providers who could not participate. Following these efforts, the median rate of Child and Family Team meetings held per active family per month doubled from program Year 1 (0.25) to program Year 3 (0.5).

Second, some Wraparound New Haven staff said they felt overwhelmed because working with a high-needs population was emotionally taxing and time-consuming. For example, to accommodate families' hectic schedules and keep them engaged, care coordinators had to meet with families before or after traditional work hours. Moreover, some care coordinators reported their caseload (average of 10 families) was difficult to manage because they provided services for all members of the family as opposed to an individual. Staff reported that finding the time needed to complete required documentation, assessments, in-home visits, and Child and Family Team meetings was a continual challenge. In addition, some staff reported that job stress was exacerbated by a lack of sufficient supervisory support, low pay, and uncertainty about whether they would keep their jobs after the cooperative agreement ended. Highlighting concerns about staff satisfaction, almost 70 percent of non-clinician staff surveyed in 2016 reported that staff retention was a major barrier (19 percent) or minor barrier (48 percent) to achieving program goals.

To help address these concerns, the awardee hired a new supervisor in program Year 3, which reduced the supervisory caseload from eight care coordinators to five care coordinators. In addition, care coordinators and behavioral health clinicians often supported and mentored one another. Most of the staff strongly (44 percent) or somewhat (48 percent) agreed that they felt supported by colleagues to do their job.

> "I think it would be helpful to have additional supports in the form of the supervisors or managers. Because although they try very hard to make themselves available, we need to have more access to supervisors. [To fill the gap], care coordinators have grown over the past year to certainly support each other and try to answer questions."
>
> —Program staff

Finally, the awardee indicated that limited community resources at times made it challenging for participants' to engage with and benefit from Wraparound New Haven. Several care coordinators reported that families had difficulty using existing transportation options (for example, because mobility impairments made bus travel difficult). Care coordinators also said it was not easy to reduce ED use given the shortage of non-emergent, after-hours care in participants' communities. In addition, changes in how Connecticut prioritized housing support services during program Year 2 made it difficult for care coordinators to resolve participants' concerns about housing insecurity.

C. Assessment of perceived program effects on the delivery of care and outcomes

In interviews, program staff felt strongly that Wraparound New Haven provided a unique resource to the community, most notably the program's holistic approach to addressing physical and behavioral health needs as well as the social service needs of the entire family. On the non-clinician staff survey, 100 percent of respondents indicated that they believed Wraparound New Haven (1) had a positive impact on participants' quality of care and services, and (2) enhanced the ability of staff to respond to the needs of families and provide services responsive to families' preferences, needs, and values in a timely way. In addition, nearly all non-clinician staff survey respondents strongly (70 percent) or somewhat (19 percent) agreed that the program was making a difference in meeting the critical needs of the community. In interviews and in the clinician survey that was fielded halfway through program Year 3, program leaders and staff conveyed similar positive perceptions of Wraparound New Haven's effect on the delivery of care.

> "I have a client that didn't know the difference about her two different [asthma medications]. And she said that she felt stupid to ask her doctor about the differences and when she should use it. I think that [participation in Wraparound New Haven] helped the client feel more comfortable with asking questions and seeing [her] doctor."
>
> —Care coordinator

Program staff said that there was anecdotal evidence that Wraparound New Haven had improved management of chronic physical and behavioral health conditions, decreased avoidable ED visits, and increased use of primary care providers:

- A care coordinator helped a family obtain a reliable source of transportation because, without it, the parent had found it difficult to maintain regular employment and get to the child's medical appointments.

- Another care coordinator noted one mother's improved medication adherence for herself and her children, including refilling prescriptions on time and maintaining medical appointments.

- One care coordinator spoke of a father who no longer used the ED as his first entry to care. The care coordinator worked with him to improve medication compliance and adherence to his treatment plan in order to be eligible for a liver transplant. His visits to the ED decreased from six times in one year to once within the past nine months.

> "I just did a Child and Family Team meeting with one of my moms and, since she started three weeks ago, she's lost 2.5 pounds. She's exercising more. The nutritionist comes in with some great ideas. She brings copies of health recipes for the family. I think it's a great add-on to the program."
>
> —Care coordinator

- Another care coordinator helped a mother of twins better understand when to use the ED, and the mother began to regularly consult the pediatrician about any concerns.

At the same time, program staff noted that even if the program demonstrated quality improvements, corresponding cost savings might not materialize during the three-year cooperative agreement.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

Program staff perceived that Wraparound New Haven had a positive impact on families' understanding of how to manage their physical and behavioral health needs. However, participants enrolled in program Year 1 and early in program Year 2—before care coordinators increased the frequency of their meetings with families or integrated physical health into the model—may have experienced fewer benefits from participation than participants who enrolled after the program improvements were made. This may have limited the impacts of the program.

In addition, even if Wraparound New Haven as designed and implemented resulted in improved health outcomes or reduced health care costs, several factors pose challenges to demonstrating those outcomes through a rigorous impact analysis of Medicaid claims data:

- Our evaluation is focused on primary enrollees and will not capture impacts of the program for enrolled family members—around a quarter of which are in the intensive physical health track.
- Many families focused on intermediate outcomes—such as losing weight or improving housing security—that were anticipated to result in long-term health outcomes. However, these impacts may not be observed within one year of the program ending (the period for detecting impacts for the HCIA R2 evaluation).

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Clifford Beers Guidance Clinic's Wraparound New Haven program.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: Clifford Beers Guidance Clinic

Evaluability domain	Response
Projected number of Medicare FFS enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018	445[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,649
Likelihood of all-cause hospitalizations	1,335
MDE sample size requirement to detect 20% effect	
Total expenditures	412
Likelihood of all-cause hospitalizations	334
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Questionable, patients may have been receiving intervention prior to HCIA R2 cooperative agreement
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	No serious issues. Proceeding with comparison group selection
Do claims identify the primary expected effects?	Some effects observed in claims data but important effects likely missing
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician and staff surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

FFS = fee-for-service.

We anticipate conducting a rigorous impact analysis of the Wraparound New Haven program. We expect to identify a pool of potential comparison beneficiaries that consists of all Medicaid beneficiaries who meet the program's eligibility criteria but live in Hartford, Connecticut (outside the program's catchment area). From this pool, we will use propensity score matching to select a comparison group that is similar to the treatment group in terms of key characteristics and prior Medicaid service use. The projected size of the analysis sample (shown in Table III.1) is based on the 445 beneficiaries who enrolled by the time enrollment ended in August 2017; the analysis sample should be sufficient to detect plausible effects on some claims-based outcomes.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Clifford Beers Guidance Clinic reached an agreement on FFS reimbursement for program services from a large commercial insurer by the end of the cooperative agreement, but was unable to reach an agreement with Connecticut's Medicaid program.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Over the course of the cooperative agreement, Clifford Beers Guidance Clinic pursued two payment models to continue funding Wraparound New Haven for children enrolled in Medicaid. First, the awardee focused on developing a value-based purchasing arrangement with the Connecticut DSS, potentially including a shared savings approach. Second, the awardee marketed Wraparound New Haven to provider organizations participating in value-based purchasing arrangements—including health systems, advanced practice networks, and independent practice associations. The awardee envisioned that provider organizations would contract with it for intensive care coordination services as a way to reduce their own overall cost of care.

C. Status of the payment model

Due to state budget shortfalls and bureaucratic hurdles, the awardee ultimately did not develop a full payment model with DSS for its original target population of children enrolled in Medicaid. Instead, Clifford Beers Guidance Clinic marketed a care coordination model based on Wraparound New Haven (renamed Advanced Care Coordination, or ACCORD) to commercial insurers. It was able to negotiate a FFS reimbursement from the state's largest insurer, Anthem Blue Cross Blue Shield. Anthem is providing reimbursement for intensive care coordination services based on the Wraparound New Haven model for commercially insured children and their families, with a focus on children with complex behavioral health needs who have not improved under current treatment plans and are at risk of hospitalization. Clifford Beers Guidance Clinic also continued to pursue a payment model agreement with the state's Medicaid program and to market Wraparound New Haven to provider organizations.

D. Factors associated with the development of the payment model

Clifford Beers Guidance Clinic faced two main barriers to developing the payment model for Wraparound New Haven: (1) state budget shortfalls and the state Medicaid agency's related hesitance to add another benefit and (2) lack of timely data on program outcomes. Given these challenges, the awardee used preliminary data on program effectiveness to market the program to providers and, in the third program year, commercial insurers. As a result of this flexible approach, Clifford Beers Guidance Clinic was able to establish a contract with Anthem for FFS reimbursement. During its no-cost extension, the awardee plans to finish its analysis of cost

savings from the program to facilitate additional discussions with payers (such as the state Medicaid program) about funding the program in the future.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The core components of the Wraparound New Haven model will not be sustained for the original focus population of children enrolled in Medicaid, but will be scaled for commercially insured children using a FFS reimbursement contract with Anthem. The awardee was also in discussions with additional commercial payers, but did not report any intent from these insurers to sustain the program as originally developed.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In program Year 2, Clifford Beers Guidance Clinic discussed funding options to sustain Wraparound New Haven with DSS, other insurers, and provider groups. To help guide these discussions, the awardee started analyzing Medicaid claims data to understand early impacts of the program on Medicaid costs. In addition, to improve the feasibility of sustaining the program, the awardee considered ways to reduce program costs after the cooperative agreement by, for example, reducing the extent of home visiting.

C. Implementing the SSR plan: progress and changes

Sustainability. As of August 2017, Clifford Beers Guidance Clinic was unable to reach an agreement with DSS to sustain Wraparound New Haven for children enrolled in Medicaid.

Scalability. Despite its inability to sustain the program for its initial target population, Clifford Beers Guidance Clinic is scaling a similar program, ACCORD, to Anthem's commercially insured population. Under ACCORD, the program will serve primary enrollees with different characteristics than those served under Wraparound New Haven. Specifically, primary enrollees will be commercially insured, the geographic scope of the program will be expanded from Greater New Haven to statewide, the age range will expand to include young adults ages 18 to 26, and youths with serious behavioral health needs may be served by the program even if they do not have a chronic physical health condition. Clifford Beers Guidance Clinic noted that it would likely need to adapt its model to accommodate these changes. For one, the awardee noted that care coordination for the families of commercially enrolled children may look different because they may face a different set of challenges and have access to different supports than Medicaid-enrolled children. In addition, Clifford Beers Guidance Clinic reported that it may need to staff hubs around the state to support the statewide expansion. To allow sufficient time for to recruit staff, the awardee was trying to phase in expansion over time.

Clifford Beers Guidance Clinic was planning to continue marketing Wraparound New Haven to additional potential funders and to use it as the model to develop programs for other populations. For example, the awardee proposed using a model similar to Wraparound New Haven to serve children with obesity through a grant with the federal Office of Minority Health.

Replicability. The awardee has developed a manual to guide others interested in replicating the program, but it has not proactively marketed the program to other clinics or states.

D. Factors associated with progress toward implementing the SSR plan

The challenges Clifford Beers Guidance Clinic faced developing its payment model with DSS—mostly importantly, state budget shortfalls—resulted in the model not being sustained for children enrolled in Medicaid. However, Anthem agreed to scale the program after the awardee shared preliminary data showing that the program had positive results on health outcomes, such as rates of depression, service utilization, and cost.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection, unless CMS requests a follow-up with the awardee to discuss the status of the payment model or its SSR plan. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Clifford Beers Guidance Clinic's Wraparound New Haven program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicaid beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Trustees of Columbia University in the City of New York

April 18, 2018

Ellen Wilson (RTI International)
Kyle Emery (RTI International)
Cordon Newhart (RTI International)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation has two interrelated goals. The first goal is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

The Trustees of Columbia University in the City of New York used funding from HCIA R2 to test the effectiveness of MySmileBuddy (MSB), an intervention in which community health workers, supported by a tablet-based software suite, worked with parents or caregivers

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The Trustees of Columbia University in the City of New York received a no-cost extension through May 31, 2018, which it is using to complete data analyses.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

(collectively referred to as "caregivers" henceforth) of young children to conduct risk assessments of the children's oral health, provide early childhood caries (ECC) management education, and develop family-level goals and action plans to prevent the progression of ECC in affected children (Table I.1). The community health workers also provided toothbrushes and toothpaste. The most prevalent chronic condition among children in the United States is dental caries, which imposes significant costs on the health care system and leads to dysfunction, disability, and distress among children. MSB's innovation was to use a nonsurgical behavioral disease management approach, as opposed to traditional surgical dental rehabilitation, to improve young children's oral health by suppressing ECC through positive home health behaviors.

The awardee was supported by multiple partners in implementing MSB, including five hospital-based pediatric dental delivery systems (PDDSs) that identified and referred eligible families and provided standard care; four community-based organizations (CBOs) that employed and supervised the community health workers, who offered oral health education and supplies to participating children; and a technical assistance provider, Health Innovation Associates, that trained and supported the community health workers. The awardee planned to enroll 1,936 children in MSB by the end of the three-year cooperative agreement.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	To improve young children's oral health by using family-level peer-counseling and behavioral risk reduction strategies supported by the use of health education technology
Major innovation	Using a chronic disease management model within pediatric dentistry to improve the care experience and reduce costs
Program components	• Patient and family engagement • Health information technology (IT)
Target population	• Children ages 2 to 6 with ECC and no comorbidities whose caregivers speak English or Spanish and are age 18 or older • Up to two eligible siblings of these children—that is, siblings in the same household younger than 6 (with caries if older than the index case or with or without caries if younger than the index case) • The parents or caregivers of these children
Theory of change/ theory of action	Columbia University hypothesizes that by educating caregivers and engaging them in goal setting (for example, cutting back on sugary drinks) and in planning how to reach these goals (for example, buying less juice), MSB will lead to changes in the caregivers' behavior that will reduce their children's ECC, thus diminishing the need for expensive surgery to treat ECC and thereby saving money and improving the children's oral health.
Payment model	Capitated payment
Award amount	$3,870,446
Effective launch date[a]	May 11, 2015
Program setting	• Recruitment conducted at PDDS clinics, Head Start day care centers and community health fairs • Services delivered at participants' homes, at locations in the community, over the telephone, or by text messaging

Table I.1 *(continued)*

Program characteristic	Description
Market area	Urban
Market location	New York City
Target outcomes	• Increased access to dental care • Improvement in the beliefs, attitudes, and self-efficacy of caregivers and community health workers • Development of caregiver-defined goals and action plans • Improvement in caregivers' self-reported use of preventive dental health behaviors with their children • Increased percentage of children who demonstrate no new cavitations beyond what existed at the time of enrollment • Reduced costs of treating ECC

[a]After the initial planning period, the awardee's program became operational as of this date.

ECC = early childhood caries; PDDS = pediatric dental delivery systems.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on four factors. First, the awardee was partially successful in its enrollment—it enrolled 1,207 participants (62 percent of its enrollment target) by the end of the initial cooperative agreement. Second, the awardee was mostly successful in hiring high quality staff and retaining them with minimal turnover. Third, MSB community health workers successfully engaged participants—making 5,401 contacts with 975 participating families, for an average of 5.5 contacts per family. Finally, participating clinicians (dental residents and their supervisors at participating PDDS sites), other implementation staff, and program participants reported that the MSB had a positive effect on participants' oral health.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Columbia University's MSB program, the analysis is still in progress and not included in this report.

Payment model. Columbia University has proposed a per beneficiary per month (PBPM) fee to be paid by a state Medicaid agency or a Medicaid managed care organization (MCO) to dentists to support community health worker services and the use of MSB software. Near the end of the third year of the cooperative agreement, the awardee was not in active negotiations with any payers but instead was focused on building an argument for a state Medicaid agency to submit a State Plan Amendment (SPA) that would allow licensed health care professionals (in this case, dentists) to delegate preventive procedures to non-licensed health workers (in this case, community health workers).

Sustainability plans. As Columbia University continued its efforts to gain approval for the payment model, it discontinued its MSB program activities. However, awardee leaders, who considered the program relatively easy to reinstate and replicate, did pursue activities to promote

the program's replication. They expected to use part of the cooperative agreement's nine-month, no-cost extension to continue planning for MSB's future.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of four data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits and written correspondence with the awardee. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on care delivery. The clinician survey, which was fielded from March 2017 to June 2017 with 71 PDDS directors and dental residents, achieved a response rate of 82 percent. The non-clinician staff survey, which was fielded from July 2016 to October 2016 with a sample of 20 MSB community health workers and program administrators, achieved a response rate of 80 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items with fewer than 11 respondents, we describe the findings in qualitative terms to avoid identifying respondents. The fourth data source was information provided by the awardee through written correspondence, using data from its administrative systems and a survey administered to program participants at the end of the program.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partly successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to enrollment strategies

Dentists identified children who were eligible for MSB during dental screenings. To be eligible, children had to be between the ages of 2 and 6, with ECC and no comorbidities, and their caregivers had to speak English or Spanish and be age 18 or older. Originally, the screening for eligible children was conducted only by dental residents from the awardee's partner PDDSs at the PDDS sites and at Head Start day care centers. During the second program year, the awardee added additional screening locations, including a pediatric medical clinic, community health fairs, and a mobile van. Although dental residents and faculty from the PDDS sites conducted the screenings at some of these locations, the awardee also hired a dentist to conduct screenings at others of the newly added locations. Once dentists identified eligible children through screening, they told the caregivers about MSB and asked them if they were interested in

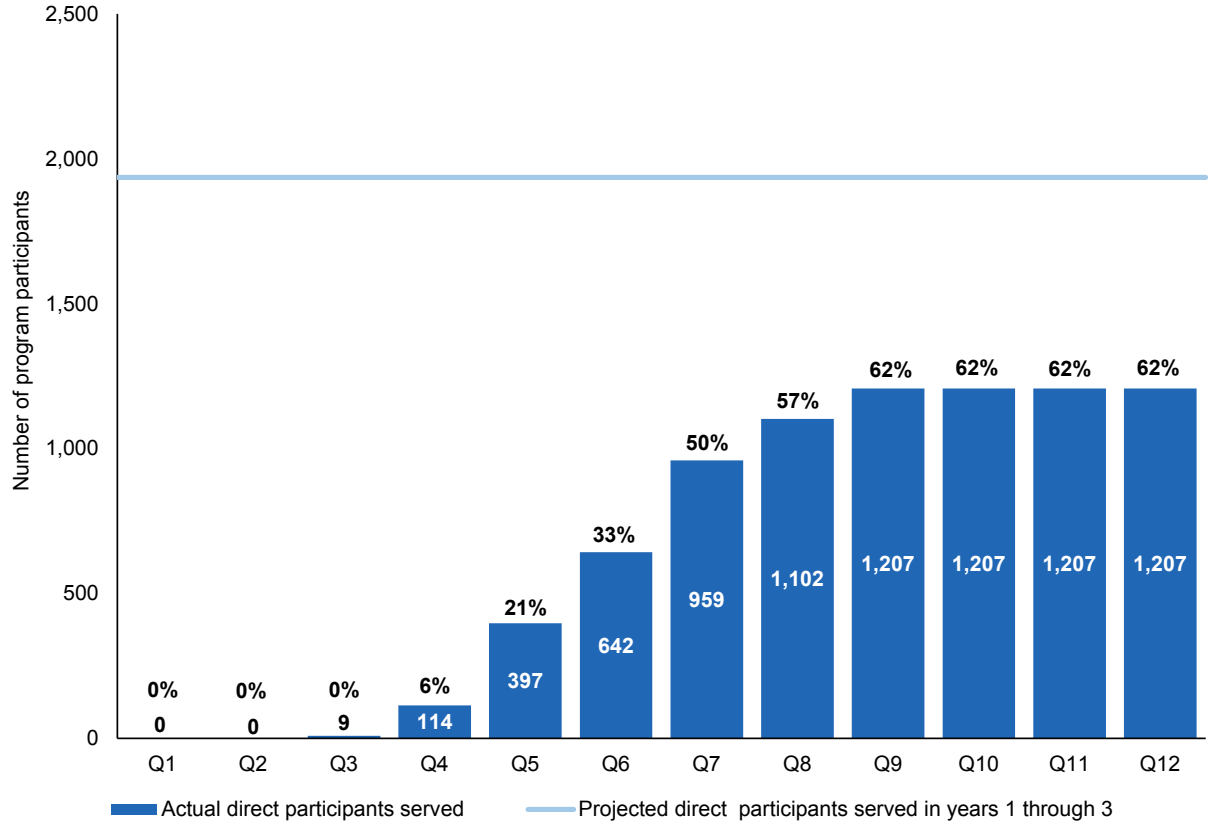
participating; if so, they asked them to fill out a form with their contact information. The forms were then collected by the awardee and distributed to the community health workers, who followed up with interested caregivers to tell them more about the program, confirm their eligibility, and enroll them if they were eligible and agreed to participate.

The awardee originally intended to complete enrollment activities 15 months before the end of the cooperative agreement (May 2016) so that all participants could receive a minimum of 12 months of the intervention before the end of the cooperative agreement. The remaining 3 months after completion of the intervention would be used for data analysis. However, because the awardee was far short of its enrollment target in May 2016, it extended enrollment by 5 months, through the end of October 2016.

b. Evidence of enrollment effectiveness

Overall, Columbia University did not achieve enrollment effectiveness. The awardee intended to enroll 1,936 direct participants for 12 months of services. By the time it stopped enrollment in October 2016 the awardee had enrolled 1,207 participants—62 percent of its projected direct participants (Figure II.1). In addition, only about half of the enrolled participants received 12 months of service. Children enrolled during the final 5 months of the enrollment period received only 6 to 11 months of services. However, in interviews, the awardee reported that 6 months of exposure to MSB was sufficient for children and families to achieve desired behavioral changes and thus potentially achieve desired oral health outcomes.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The enrollment projection of 1,207 submitted by Columbia to the implementation and monitoring contractor in the ninth program quarter is lower than the target of 1,936 reported by the awardee in our July 2017 site visit interviews.

c. Barriers and facilitators associated with enrollment effectiveness

Based on interviews with program staff and partners, Columbia University's progress in meeting its three-year enrollment goal was influenced by several factors: (1) fewer PDDS sites agreeing to participate in MSB than anticipated, (2) a delay in the start-up of the program, (3) fewer referrals from participating dentists than expected, and (4) some caregivers' lack of interest in participating. These problems emerged at the beginning of the first program year and persisted to varying degrees throughout the enrollment period (which ended early in the third program year), despite the awardee's attempts to address the problems.

First, Columbia University had originally planned to partner with 12 PDDSs to identify eligible participants. However, after receipt of the HCIA R2 cooperative agreement, it found that many of the PDDSs were not willing to commit to the program's disease management approach.

As a result, the awardee partnered with only 5 PDDSs, which reduced the number of children screened.

Second, the awardee initiated enrollment activities two and a half months later than originally planned. A variety of challenges contributed to the delay, including the need for unexpected updates to the MSB software suite, a university requirement to develop a custom data management system because of a data breach on a different project, and delays in receiving institutional review board approvals from participating PDDSs and CBOs.

Third, the program received a lower-than-expected number of referrals from the Head Start day care centers and dentists at the five partner PDDS sites. Reasons for the low number of referrals included (1) fewer children than expected at the Head Start day care centers had caries, so fewer were eligible for the program; (2) dental residents doing the screenings at the PDDSs had many demands on their time and often did not have time to, or forgot to, talk to caregivers about the program; and (3) PDDS staff, including the dental residents, did not fully buy into the program because financial and training incentives at the PDDSs were aligned with conducting procedures, not prevention. The awardee had a subcontract with New York University (NYU) to coordinate and oversee screening at the PDDSs and Head Start day care centers. However, by the middle of the second program year, the awardee had concluded that NYU was not effective in addressing the recruitment challenges and dissolved its subcontract.

> "In pediatric dental residency programs, the residents are judged by numbers of procedures [and] the program directors are judged by dollars based on procedures."
>
> —Program leader

During the second program year, the awardee took several steps to try to overcome these barriers and increase the number of referrals to the program. For example, it expanded the program's eligibility criteria to include certain siblings of eligible children. It also expanded the number of locations where it screened children for eligibility by adding a pediatric medical clinic, a mobile van, and community health fairs as screening sites. The awardee and its partners also tried numerous strategies to motivate dental residents to discuss the program with eligible families, such as talking about the importance of making MSB referrals during their huddles, highlighting the names of children who fell within the eligible age range on the daily printed appointments schedule, and posting flyers about MSB where dental residents and families could see them. The community health workers also started spending more time at the PDDS sites. Their presence helped remind dental residents of the program and provided the community health workers an opportunity to speak to families directly about the program. To increase residents' motivation to refer children to MSB in the latter half of the second program year, the principal investigator gave a series of presentations to PDDS faculty and residents and started holding monthly webinars. The presentations and webinars were intended to help residents understand the importance of a disease management approach to improving children's oral health and to make the case that the field of dentistry was changing and the disease management approach was likely to be the model of care for most of their careers. The awardee also disseminated a newsletter for dental residents that kept them up-to-date on the program and promoted competition among PDDS sites by presenting referral numbers by site. During interviews in the second program year, a resident said that these efforts had been effective in increasing awareness and motivation to refer families to the program. Enrollment numbers, however, did not significantly increase.

Interviewees identified strategies that, if they could start the intervention over again, might have increased the rate of referrals to the program. For example, staff at one PDDS said that, although modifying the electronic charting system would have been complicated, including a check box for MSB in the system likely would have increased referral compliance among residents. While reflecting on enrollment efforts, the awardee's principal investigator said he wished that he had engaged in direct conversations with PDDS directors and their leaders (typically, hospital executives) before initiating the project to ensure that they were committed to MSB—even if participation meant that dental residents would perform fewer procedures than was typical. He also suggested that a program like MSB might be more effective if conducted in pediatric medical settings instead of dental settings because medical providers are more familiar with the disease management model.

The fourth barrier to enrollment described by awardee leaders, PDDS staff, and community health workers was that some eligible families did not consider their children's oral health to be a salient issue and were therefore reluctant or not motivated to enroll in the program. According to interviewees, many eligible families faced multiple challenges in their lives that took precedence over MSB. In addition, most did not have a disease management orientation to their children's oral health. Instead, they typically felt that they were providing their children with appropriate care as long as they took their child to the dentist periodically to get any cavities filled. MSB community health workers and program administrators who completed the non-clinician staff survey reported that the main reasons eligible families may not participate in MSB were that the services seemed unnecessary to them (81 percent), the program required too much of a time commitment (69 percent), they did not want to be enrolled in a research study or government-funded program (63 percent), and they did not understand the program (56 percent). Dental residents who conducted screenings and supervisors who completed the clinician survey generally had similar responses. They reported that the primary reasons families did not participate were that there were too many requirements for participants (62 percent), they did not understand the program (55 percent), and it was too much of a time commitment for families (53 percent).

Although the community health workers tried to help the families understand the value of oral health prevention and make participation in the program as easy as possible (for example, being flexible about meeting times and locations to accommodate caregivers' schedules), they were often unable to overcome families' reticence to enroll in MSB. Program leaders and a PDDS director also commented that some families did not like the idea of a community health worker coming to their home because it felt intrusive. To overcome this barrier to enrollment, the awardee modified its protocols for family meetings so that the community health worker could meet with the family not only at the family's home, but also at the PDDS or any public space that was comfortable for them.

2. Delivery of program services

a. Description of and changes to the service delivery model

Columbia University had previously received a series of grants from the National Institutes of Health (NIH) to develop, validate, and field test the MSB software suite. The awardee viewed the HCIA R2 cooperative agreement as an effectiveness demonstration, expanding on its previous work to test implementation of MSB on a much larger scale and in the community. For

the cooperative agreement, the awardee (1) established relationships with PDDS partners and developed the processes to recruit participants, (2) established relationships with CBO partners that hired and supervised the community health workers, and (3) developed the protocols for community health workers to meet with families and provide oral health education and supplies during multiple encounters over 12 months.[4] The awardee had expected that the MSB software suite would not require any changes, but it found that changes had occurred to the software platform on which MSB was originally developed and that extensive software updates were required to effectively host MSB.

Key steps in MSB service delivery model included having community health workers perform the following duties:

- Schedule initial meetings with caregivers of eligible children in a location that is comfortable for the family.
- Enroll families during the initial meeting.
- Use MSB software to conduct an ECC risk assessment, provide caregiver education and assistance, compute an individualized risk score, set family goals, and develop an action plan with the family during the initial meeting.
- Conduct follow-up meetings (in person or by telephone) with families to assess progress toward their goals and troubleshoot any problems they may be facing in implementing their action plan. The frequency of the meetings depends upon the families' needs and availability, but community health workers should attempt to meet with the families in person at least once every three months to provide new toothbrushes and toothpaste.

If these steps were followed, the awardee hypothesized that caregivers would take the steps to improve their children's oral health (for example, giving them fewer sugary beverages and brushing their teeth more frequently). These behavior changes would then arrest the ECC disease process. As a result, children would require fewer dental repairs, which would cause overall dental repair costs to decrease.

b. Evidence of service delivery effectiveness

Overall, Columbia University was effective in its implementation of MSB program. The awardee successfully provided core MSB services, including education and support for behavior change to program participants; staffed most of its positions and trained the staff; and engaged participants. However, the awardee faced some difficulties with hiring, retention, and maintaining adequate supervision for the community health workers.

Delivery of intervention services. According to program leaders, PDDS staff, and community health worker supervisors, the awardee was largely successful in delivering MSB services to children and their families as intended. They said, for example, that the community health workers effectively communicated with families and that, with the support of MSB software suite, they were able to conduct risk assessments, develop action plans, and provide detailed and accurate oral health prevention information and oral health supplies to families. The

[4] In its previous tests of the MSB software suite, community health workers had met with families just one time.

awardee reported that between May 2015 and April 2017 the 11 community health workers engaged in more than 5,000 successful encounters with the 975 families enrolled in the program through telephone, text, and home and community-based visits—an average of 5.5 contacts per family. The awardee also reported that 85 percent of these encounters included discussions of dental care (for example, inquiry about dental care, encouragement to seek dental care, facilitation of access to dental care) and 10 percent involved the community health worker attending a dental visit together with the child and caregiver. In addition, 14 percent of encounters dealt specifically with overcoming non-oral health issues that constituted barriers to oral health and dental care (for example, food, income, and housing insecurities; language barriers; immigration and legal issues; and domestic violence). Although the awardee had originally planned for each community health worker to conduct six to eight contacts per family over the course of their 12-month participation, it later relaxed that requirement because it found that some families did not need as many as six contacts.

Staffing and training. Columbia University had mixed success in hiring, training, and retaining staff. Among the program's administrative staff, the awardee had no turnover during the three years of the program. For the community health workers, however, the awardee experienced hiring delays during the first program year due primarily to issues related to union rules. Delays in community health worker hiring meant that training had to be repeated several times as community health workers were gradually brought on board. Retention of the community health workers was fairly good. Although one community health worker resigned just weeks after being hired in the first program year, no additional community health workers left the program until the third program year. During the third and final program year, turnover was higher because one partner CBO laid off a number of its staff, including two community health workers, and because Columbia University dismissed one community health worker for poor performance. In addition, some of the CBOs experienced significant turnover among the community health workers' supervisors during the course of the three-year cooperative agreement. In fact, the awardee dissolved its relationship with one of its partner CBOs in the third program year because of frequent turnover in the supervisor role and a lack of support to the community health workers based there. The awardee then hired the CBO's community health workers as independent contractors and supervised them directly.

In interviews, community health workers and their supervisors said that the training they received at the start of the program was excellent. All community health workers and supervisors who completed the non-clinician staff survey strongly agreed that the training they received helped them improve their job performance and that they learned new skills that were important for their job. In interviews, community health workers reported that the biweekly meetings held by the awardee were a valuable opportunity for them to support each other and brainstorm ways to address challenges they encountered.

Recruitment and engagement of providers. With the exception of educating dental providers on the benefits of nonsurgical disease management for children's oral health care and encouraging them to refer participants to MSB, Columbia University did not include direct engagement of providers as a core component of its service delivery model.

Engagement of program participants. Columbia University was largely successful in engaging participants. In interviews, program leaders and community health workers said that

most participants attended scheduled encounters and participated actively in MSB activities, although a small number of families formally withdrew from the program and some were lost to follow-up because they did not respond to repeated attempts to contact them. In addition, participants who completed the end-of-program survey conducted by the awardee reported that they had high levels of satisfaction with MSB. Among the 650 caregivers who responded to the awardee's survey, 99 percent agreed with the statement, "I would recommend the MySmileBuddy Program."

c. Barriers and facilitators associated with service delivery effectiveness

Columbia University's ability to effectively deliver MSB intervention services was influenced by several factors. Key facilitators of service delivery were (1) having community health workers who were from the community and were passionate about their work and (2) the MSB software suite. Barriers to service delivery effectiveness included (1) lack of support for community health workers at some CBOs and (2) the complexities of families' life circumstances.

Across all three program years, interviewees stated that the main factor that contributed to MSB's effectiveness in delivering program services was the use of community health workers who were passionate about their work, who identified with the participants' culture, and who lived in participants' communities. These traits enabled the community health workers to understand participants and build trust with them, which in turn kept the families engaged and helped them overcome challenges to improving their children's oral health. In the end-of-program survey conducted by the awardee, participants reported high levels of satisfaction with the community health workers' accessibility, accountability, and engagement. For example, 97 percent or more of participants reported that their community health worker was warm and friendly, explained things in a way they could understand, and helped them reach their goals.

As described in the first HCIA annual report,[5] the MSB software suite facilitated service delivery. The software supported the community health workers in providing accurate oral health information to families because it included the preventive oral health information that community health workers needed to convey to participants. It also included templates for conducting risk assessments and developing action plans. The software also helped to engage participants. As one interviewee noted, the software was designed for caregivers and children to be able to use on their own: it was cartoony, fun, and interactive. In the end-of-program survey conducted by the awardee, caregivers rated MSB software suite highly, with 98 percent or more of respondents rating the software as easy to understand, visually appealing, and helpful for reaching their oral health goals.

[5] See https://downloads.cms.gov/files/cmmi/hcia2-yroneevalrpt.pdf.

> "One of our less successful areas had to be the constant turnover of [community health worker] supervisors, which is dispiriting to a [community health worker]. They have to keep trying to tell their [new] supervisor what they're doing."
>
> —Awardee leader

The turnover of supervisors at some of the partner CBOs, however, hindered the awardee's ability to deliver intervention services. According to program leaders, community health workers at sites with supervisor turnover did not receive the consistent guidance and support they needed to effectively deliver MSB services. In the second program year, the awardee's principal investigator said that if he had to do it over again, he would hire community health workers directly, rather than basing them in partner CBOs.

Another challenge to service delivery effectiveness was the complex life circumstances of many families (for example, food insecurity, low incomes, inadequate or unstable housing, speaking English as a second language, domestic violence, and immigration and legal issues), which made enrolling in and remaining in the program difficult. In the third program year, interviewees also noted a new barrier to families' participation: the anti-immigrant political atmosphere under the new presidential administration. Because most of the MSB participants were Spanish-speaking immigrants, community health workers said that this atmosphere made them less willing to engage with the program, both because they were afraid to share their personal information with a government-funded program and because they were afraid to present at the dentist's office or at the community health workers' offices for fear of being deported. Community health workers worked to overcome these challenges by helping families address the non-oral health challenges in their lives as best they could, being flexible when scheduling follow-ups, and assuring them that their participation in the program would not increase their risk for deportation. According to awardee leaders, one lesson learned was the importance of allowing community health workers flexibility in terms of the timing, quantity, and types of support they provided to families and the strategies they used to communicate with them. Given that the families differed in terms of the amount of support they needed and the challenges they encountered, the awardee found that community health workers were best positioned to tailor the program to each family's needs.

C. Assessment of perceived program effects on the delivery of care and outcomes

Most PDDS dental residents and their supervisors, MSB staff, and community health workers and their supervisors believed that the program had strong positive effects on program participants and their oral health beliefs and behaviors. In interviews, community health workers, their supervisors, and program administrators recounted anecdotal evidence of significant behavioral change in families. They said that caregivers had, for example, started putting water instead of juice in their children's sippy cups or started brushing their children's teeth twice a day. A community health worker supervisor and a program administrator also cited anecdotal evidence that MSB families were more likely to attend recommended dental care visits and that when families

> "When the community health workers initially met with the family, they did not have regular dental brushing practices whatsoever or they would put juice in the baby bottle and constantly feed their children candy. But … at the final [MSB] visit, there was a great difference in the quality of the child's teeth and many mothers had instilled new practices in their families."
>
> —Community health worker supervisor

returned to see dental providers their oral health had often improved. In addition, the majority of MSB administrator and community health workers (88 percent) and dental residents and their supervisors (86 percent) who completed the non-clinician and clinician surveys reported that MSB had been very or somewhat effective in achieving its goals.

Finally, participants who completed the survey conducted by the awardee both before and after program participation reported that their children's oral health had significantly improved. For example, before participating in MSB, 51 percent of caregivers rated their child to be in poor or fair oral health. After participating in MSB, only 2 percent of caregivers rated their child's oral health as poor or fair.

D. Implications of implementation findings for the design and interpretation of an impact analysis

In this section, we highlight the implications of our findings for the design and interpretation of an impact analysis.

As implemented, the awardee did not achieve its full enrollment target or its originally intended length of service exposure for program participants. Although the awardee fell short of its enrollment target, the final number of enrolled participants was large enough that an impact analysis likely would have enough statistical power to detect moderately sized effects of the program. In addition, although many participants received only 6 to 11 months of services (rather than 12 months as originally planned), the awardee ultimately determined that 6 months of exposure was sufficient to change caregivers' behaviors related to their children's oral health and that these changes in behaviors would lead to improvements in the children's oral health. Therefore, we would not expect that the reduced length of service exposure would reduce the likelihood of uncovering positive program impacts on either caregiver behaviors or children's oral health in an impact analysis.

Although interview and survey responses from community health workers, awardee leaders, and clinicians involved with MSB program support the finding that the awardee was partly successful in implementing the program, any improvements in children's oral health may not be detectable in an impact evaluation, for several reasons. First, hospitals frequently do not file claims for dental procedures, so Medicaid claims data may not be a reliable indicator of any changes in the number of oral health–related procedures performed. Second, dental claims do not include diagnosis codes, so their value in monitoring ECC experiences is somewhat limited. Third, the awardee commented that dentists may be inclined to repair teeth, even if the disease process has been halted and repair is not necessary. As previously noted, dental residents at PDDSs have strong incentives to perform procedures. In addition, many dentists believe that damaged primary teeth need to be repaired, even if the disease process has been arrested. Thus, even if the intervention were successful in improving participants' oral health behaviors and oral health, improvements may not lead to a reduction in the number of oral health procedures carried out. If the number of procedures implemented were unaffected, oral health care costs would also likely be unaffected.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Columbia University's MSB program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Columbia University

Evaluability domain	Response
Projected number of Medicare fee-for service (FFS) enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Projected Medicaid population with 6 months of program exposure by February 28, 2018	846[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	2,932
Likelihood of all-cause hospitalizations	2,219
MDE sample size requirement to detect 20% effect	
Total expenditures	733
Likelihood of all-cause hospitalizations	555
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Questionable, no testing yet to determine strength of intent-to-treat framework
Likelihood of solid comparison group	Too early to determine due to delay in Medicaid data
Do claims identify the primary expected effects?	Some effects observed in claims data but important effects likely missing
Core outcomes estimation method	Possible difference in differences
Primary reason for no rigorous evaluation	Rigorous analysis will be possible if a large percentage of treatment group received dental care prior to intervention
Survey data for treatment group that will be analyzed	Clinician and staff surveys
Implementation data that will be analyzed	Awardee is analyzing its clinical data

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We anticipate being able to use propensity score matching to select comparison beneficiaries who had similar characteristics and dental service use as treatment beneficiaries did during the baseline period. However, to implement this strategy, we will need to limit the treatment group to those who received dental care during the baseline period. Because of lags in Medicaid data, we do not know how many of the program's 846 enrollees shown in Table III.1 (as of August 2017) actually received dental care prior to the intervention, so we are not certain that our sample size will be sufficient. Although it is unclear whether some of the key outcomes that the awardee is trying to affect (such as dental surgeries) will be observable in the claims data, at least some of the awardee's outcomes (for example, receipt of dental care and total expenditures) will be captured in the claims data.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Columbia University has proposed a PBPM fee to be paid by a state Medicaid agency or a Medicaid MCO to dentists to support community health worker services and the use of MSB software. Near the end of the third year of the cooperative agreement, the awardee was not in active negotiations with any payers, but was focused on building an argument for a state Medicaid agency to submit an SPA that would allow licensed health care professionals (in this case, dentists) to delegate preventive procedures to non-licensed health workers (the community health workers).

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leader who was most familiar with the payment model.

B. Description of the payment model

Columbia University's proposed payment model consisted of a PBPM fee, paid by a state Medicaid agency or a Medicaid MCO that included or was focused on dental care, for dentists to support community health worker services and the use of the MSB software suite for children at high risk of dental caries or who already have caries. According to the awardee, the payment model was based upon Medicaid's Primary Care Case Management model.

Under the proposed model, payment would go from Medicaid (or an MCO) to the dentist, who would pay for the MSB software suite and for a community health worker to provide the preventive care. The services provided under the payment model would include education, risk assessment, goal setting, creation of action plans, facilitation, and follow-up. A dentist would identify children who were suitable for the program and refer them to the intervention. The intent of the payment model would be to enroll children who are at the highest risk of caries or caries progression. The high-risk children could be identified by using a standard dental examination or a test of the level of streptococcus mutans in the child's saliva.[6]

The amount of the PBPM has not been determined. According to the awardee, it had enough information from the HCIA R2 demonstration about the services required to implement the intervention (for example, how much time is needed from community health workers) to calculate the PBPM near the end of the third year of the cooperative agreement. The awardee anticipated that the payment model would likely need to pay for community health worker services for up to nine months for families that had underlying social challenges (for example, inadequate housing, domestic violence, food insecurity) and four to six months for families without underlying social challenges.

[6] See Edelstein, B. L., S. D. Ureles, and A. Smaldone. "Very High Salivary *Streptococcus Mutans* Predicts Caries Progression in Young Children." *Pediatric Dentistry,* vol. 38, no. 4, 2016, pp. 325–330.

The awardee noted that because the New York State Medicaid Redesign team was opposed to any kind of fee-for-service (FFS) payment system, a payment model for the program in New York would need to use a value-based purchasing (VBP) approach—for example, a bundled payment option that rewards improvements in quality and outcomes. The awardee identified several outcomes that could be suitable measures for a VBP model, including behavioral changes (for example, caregiver-reported adoption of brushing twice daily with fluoridated toothpaste) or clinical outcomes (for example, condition of teeth as revealed by a dental examination or the level of streptococcus mutans in saliva). The awardee speculated that other states might be willing to consider a payment model that would allow community health workers to be paid directly on a FFS basis (for example, per visit with a family).

C. Status of the payment model

Near the end of the third year of the cooperative agreement, the awardee was not in active negotiations with any payers. Its main focus was to convince a state Medicaid agency to submit an SPA that would allow licensed health care professionals (the dentists) to delegate preventive procedures to non-licensed health workers (the community health workers). The awardee engaged a consultant to support its efforts by reviewing the relevant regulations, developing (if appropriate) a model SPA, working with the awardee to vet the model SPA with CMS and one or more states, and facilitating its adoption (at least as a pilot) in one or more states.

The awardee was also engaged in discussions with a group of organizations in New York that were interested in having the state submit an SPA that would allow Medicaid to support the work of community health workers more broadly. These partners included the New York City Department of Health and Mental Hygiene (another HCIA R2 awardee), Community Health Worker Network of New York City, and the Schuyler Center for Analysis and Advocacy. Some in the group wanted a broad SPA that would apply to all of the activities that community health workers do generally (regardless of a specific disease being addressed), whereas Columbia University's interest was primarily in using an SPA for the specific dental health–related community health worker services being provided as part of its intervention.

D. Factors associated with the development of the payment model

The awardee described four major challenges to the development of a payment model for the program: (1) limited control over the activities of subcontractors under the current program structure, (2) conflicts with the current overall organization of dental care delivery, (3) conflicts with the current organization of dental care payment mechanisms, and (4) intellectual property considerations.

First, the awardee reported that limited control over the activities of subcontractors (for example, hiring and training the community health workers or delivering direct dental services) such as during the award would make success under the proposed payment model difficult. The awardee believed its payment model could best be implemented within a single accountable care organization, patient-centered medical home, or other unified delivery system. That way, all parties involved (dentists, pediatric primary care medical personnel, dental trainees, program recruiters, program directors and faculty, and community health workers) would work for, be paid by, and be accountable to a single institution.

Second, dental clinicians are resistant to the idea of nonsurgical disease management for dental caries as an appropriate form of treatment because the financial incentives and the culture and training of dentistry have not historically supported it. The payment model could help address the barrier of the financial incentives, but dentists still may be reluctant to adopt the intervention and engage with the payment model because they do not have a disease management orientation.

Third, the rigidity of dental payment mechanisms could make it difficult to implement a new payment model even when the parties involved are motivated to change the approach to care. For example, for another project, the awardee is working with a Medicaid MCO, a hospital, and a Medicaid authority to implement a pharmacological caries management intervention. According to the awardee leaders, they are having difficulty implementing the payment model because paying for a range of services for disease management instead of paying for specific services would require fundamental changes in the MCO's software and approach to adjudicating claims.

Fourth, a further challenge with the payment model is the fact that MSB software suite is university-owned intellectual property. For the awardee leaders to be able to share it with other interested parties, they would have to form a business that could license the technology.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

As Columbia University continued its efforts to gain approval for the payment model, it discontinued MSB program activities. Program leaders noted, however, that several of the PDDS sites were continuing to educate residents about disease management. The awardee considered the program relatively easy to reinstate and replicate and pursued activities to promote the program's replication elsewhere. The awardee intended to use part of its nine-month, no-cost extension to continue planning for the program's future.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, Columbia University had minimally developed plans for sustaining its MSB program. The awardee had received a one-year planning grant from NIH to conduct an efficacy trial and had applied for the next stage of funding. At that time, the awardee had no plans to scale the program, but was hopeful that others might if they had financial incentives to promote disease management. At the time, the awardee was working with a group in California that had access to funding through the state's Medicaid waiver. An entity in Alaska was interested in partnering with Columbia University to replicate the program if they could secure grant funding.

The awardee expected further replication of the program if dental payments moved to pay-for-value arrangements.

C. Implementing the SSR plan: progress and changes

Sustainability. In Year 3, Columbia University learned that its NIH application to conduct an efficacy trial of MSB was not funded. The awardee was trying to gain support for sustaining the MSB program through third-party payments that support the community health worker functions. The awardee was in discussions with the federal and state Medicaid agency about a potential SPA to develop a way to pay for community health worker services (via a licensed professional, such as a dentist), but did not yet have these in place. As another tool to try to obtain more state support for the payment model, the awardee commissioned a documentary film about the program to highlight the work of the community health workers.

The awardee stopped implementing MSB midway through the third year because it did not have the funds to continue the program. Still, Make the Road, one of the four CBOs that oversaw community health workers for the program, was trying to find ways to sustain MSB activities. Make the Road staff planned to attend the Oral Health Coalition Regional Conference to brainstorm with other team-based organizations and providers in the dental field on potential funding opportunities for the program. The community health workers no longer had tablets (they belonged to Columbia University), but could instead use laptops to administer the program.

Even after MSB had stopped program activities, however, some of the PDDS sites continued educating pediatric dental residents about the benefits of a disease management approach to oral health. The education of dental residents was an activity the awardee conducted to boost enrollment, however, not a core component of the MSB program.

Columbia University received a nine-month, no-cost extension to complete data analyses and reporting and further its plan for sustainability.

Scalability. The awardee made no efforts to scale MSB program to more patients in its service area during the cooperative agreement.

Replicability. MSB has not yet been replicated elsewhere, but the awardee pursued activities to help foster replication. It published a paper in *Dental Clinics of North America* that explained why a medical approach to oral health was better than a surgical one and why partnering with community health workers and assisting dental professionals was important for implementing a medical approach. In addition, the awardee was involved in efforts of the Innovation Accelerator Program at the Centers for Medicaid and CHIP Services (CMCS) to identify childhood caries management programs across the country with sufficient evidence to be potential models for other states to replicate. CMCS selected MSB as one of the 13 models for states to apply for technical assistance to implement.

D. Factors associated with progress toward implementing the SSR plan

The awardee thought that the country may have reached a tipping point in recognizing the need to better prevent dental caries and manage dental disease for children. The American Association of Pediatric Dentistry issued guidelines for treating ECC that align with MSB

program. The awardee was hopeful that if a large state (such as California, Florida, Michigan, or New York) achieved a Medicaid SPA to authorize community health workers to receive direct payment for ECC management, it could potentially prompt organized dentistry, dental insurers, and child advocates to implement ECC management more broadly.

The awardee reported that, because MSB had been refined over nine years, it was ready for SSR. The awardee also thought the design of the program facilitated SSR, describing it as a "turnkey, implementation-ready program" supported by protocols and technology that required minimal training. The program team reportedly could train community health workers for this program in a single day. As one awardee leader reported: "They'd be off and running. All they need is somebody to pay them."

Indeed, lack of ongoing payment for the community health workers was the key barrier to SSR. In addition, in trying to get payer support for the payment model, the awardee found it difficult to have open conversations and make progress because of competition among dental insurers around disease management approaches. In reflecting on discussions with an insurer, the awardee stated: "The good news is he felt that there is energy in the industry to fund what we've got to offer. The bad news is that everybody wants to consider it proprietary." Finally, one respondent identified a secondary barrier to SSR as the need to continually update the health IT components of the program.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Columbia University's MSB program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicaid beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Detroit Medical Center

April 18, 2018

Margaret O'Brien-Strain (Mission Analytics Group, Inc.)

Smita Patil (Mission Analytics Group, Inc.)

James Reschovsky (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Detroit Medical Center received a no-cost extension from CMMI to extend the Gateway program through December 31, 2017. Detroit Medical Center stopped enrolling Gateway participants that are paid for by HCIA R2 funding on August 30, 2017. However, Detroit Medical Center will continue to operate its program through December 31, 2018 through funding from Tenet Health, its parent organization. Detroit Medical Center will continue to enroll participants on an ongoing basis and will pay for all Gateway participants enrolled after August 30, 2017 with funding from Tenet Health.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Detroit Medical Center is using funding from HCIA R2 to support the Gateway to Health program. Detroit Medical Center designed the Gateway program to provide ongoing primary care services in a patient-centered medical home (PCMH) model to people living in Detroit, Michigan. The primary target population includes individuals who have been identified as frequent users of the emergency department (ED), have no primary care physician (PCP) on record, and have at least one of the following chronic conditions: (1) diabetes, (2) asthma, (3) hypertension, (4) congestive heart failure (CHF), (5) depression, (6) chronic obstructive pulmonary disease (COPD), or (7) HIV/AIDS. In addition, Detroit Medical Center has established partnerships with various organizations to solicit referrals to the Gateway program.

Once enrolled in the Gateway program, participants receive ongoing primary care services from a multidisciplinary care team. The care team is led by a clinical provider—either a nurse practitioner (NP) or PCP—and comprises a behavior health specialist, social worker, nutritionist, pharmacy educator, and a medical assistant. The multidisciplinary nature of the Gateway care teams allows Detroit Medical Center to identify and address the full range of participants' physical, behavioral, and social needs.

The Gateway centers are located within or adjacent to EDs operated by Detroit Medical Center. The colocation of the Gateway clinics with EDs is precisely what makes the program innovative. Detroit Medical Center leaders believe that increasing Gateway participants' access to a usual source of primary care will lead to better management of chronic conditions and fewer ED visits.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The Detroit Medical Center (DMC) designed the Gateway program to provide ongoing primary care services in a PCMH model to frequent ED users who have no provider and one of seven chronic conditions.
Major innovation	Colocation of PCMH primary care clinics with EDs
Program components	• Medical home • Education and training
Target population	People living in Detroit who have been identified as frequent users of the ED who have no PCP and at least one of the following chronic conditions: (1) diabetes, (2) asthma, (3) hypertension, (4) CHF, (5) depression, (6) COPD, and (7) HIV/AIDS
Theory of change/ theory of action	DMC focuses on changing participants' reliance on the ED for medical care by offering participants who seek treatment there the option of receiving immediate access to primary care at a Gateway center. This improved access to primary care will result in better health outcomes, fewer ED visits, and lower costs.
Payment model	Capitated payment, shared savings, fee-for-service (FFS)
Award amount	$9,987,542

Table I.1 *(continued)*

Program characteristic	Description
Effective launch date	January 20, 2015
Program setting	Gateway centers located at or adjacent to three of Detroit's largest EDs
Market area	Urban
Market location	Detroit, MI
Target outcomes	• Increase PCMH use • Decrease ED use and service costs • Improve overall health among target patients

CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; ED = emergency department; PCMH = patient-centered medical home.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on four factors. First, the awardee enrolled 6,996 participants—85.3 percent of its enrollment target—by the end of the initial cooperative agreement. Second, the awardee was able to deliver services to program participants through the Gateway care model as planned, without any significant changes. Third, the awardee experienced only periodic staff turnover throughout the cooperative agreement and experienced high rates of staff satisfaction. Fourth, the awardee was able to engage its program participants and, according to awardee staff, appears to have positively impacted participant health outcomes. Finally, participating implementation staff reported that the program had a positive effect on the delivery of care.

Impact evaluation. Due to the lack of a strong comparison group, we do not anticipate being able to conduct a rigorous impact analysis for Detroit Medical Center. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. Detroit Medical Center planned to support its program through existing fee-for-service (FFS) billing after the cooperative agreement and did not propose a new payment model.

Sustainability plans. Detroit Medical Center expected to sustain its Gateway program at two of its three original sites after the cooperative agreement ended, through internal support and FFS payments. Detroit Medical Center's parent organization, Tenet Health, agreed to provide support funding to continue the program for another year at two of the three original sites. The awardee has tried to secure relationships with other providers—both primary care and hospitals —to sustain and potentially scale the program throughout the city, but has struggled to demonstrate how the program directly helps other hospitals.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a non-clinician staff survey on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July 2016 to October of 2016 with a sample of 30 potential respondents and achieved a response rate of 71 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

The awardee primarily recruited and enrolled participants when they presented in the ED. Detroit Medical Center tagged patients who met the Gateway program eligibility criteria in the electronic medical record (EMR) system as "PG" (for Potential Gateway). When a PG patient presented in the ED, a triage nurse conducted a uniform triage assessment through which the patient was assigned a nationally standardized Emergency Severity Index Score on a scale of 1 (highest acuity) to 5 (lowest acuity). Patients with a score of either 4 or 5 were considered low acuity; they underwent a medical screening examination (MSE) conducted by a triage nurse. Depending upon the results, the triage nurse connected the patient to a Gateway navigator. The navigator asked the patient whether he or she had a PCP. If the patient did not have a PCP, the navigator offered the patient an opportunity to receive care at a Gateway center. The navigator

provided a pamphlet explaining the Gateway program, the types of services provided by the center, the patient's option to receive care through a PCMH, and the freedom to opt out at any time. Navigators escorted consenting patients to the appropriate Gateway center.

Eligible patients who present in the ED with a true medical emergency are appropriately treated in the ED instead of being referred to a Gateway navigator. In these cases, navigators (if the patient is treated in the ED) or social workers (if the patient is admitted to the hospital) inform the patients that they may receive follow-up services at a Gateway center after they are discharged. Gateway program leaders helped create an electronic order for a Gateway consultation, which can be placed through the EMR by a clinician treating an admitted eligible patient. This electronic consult triggers a Gateway social worker to visit the admitted patient at bedside to enroll the patient in the Gateway program.

In the second program year, Detroit Medical Center expanded the Gateway program eligibility criteria to allow current participants to refer other potentially eligible individuals. For example, the awardee expanded the eligibility criteria to include Detroit Medical Center employees. Gateway program leaders partnered with the Detroit Medical Center occupational health department to create a pathway to enrollment for Detroit Medical Center employees. Detroit Medical Center employees that meet the Gateway program eligibility criteria receive information regarding the program from the occupational health department. Eligible employees can then self-refer for enrollment to the Gateway program. Detroit Medical Center also expanded the program eligibility criteria to include vulnerable populations, such as nursing facility residents and foster youth. In addition, Detroit Medical Center established partnerships with external organizations to allow for those organizations to refer individuals who meet the program eligibility criteria for enrollment in the Gateway program.

Detroit Medical Center enrolled participants into the Gateway program between January 20, 2015, and September 1, 2017. Tenet Health, the parent organization of Detroit Medical Center has agreed to fund the Gateway program through December 2018. Thus, the enrollment period gave participants ample time for exposure to intervention services.

b. Evidence of enrollment effectiveness

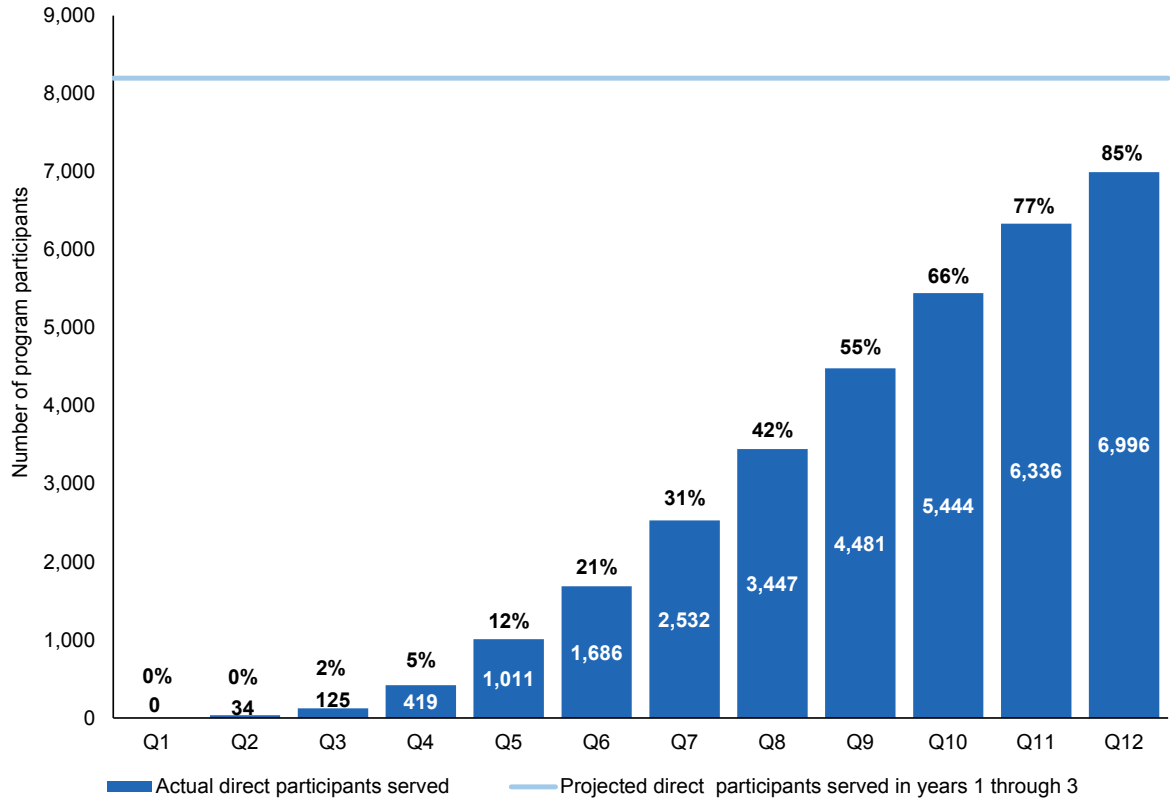
Detroit Medical Center partially achieved enrollment effectiveness. The awardee enrolled 6,996 participants from January 2015 (when it launched its program) through August 2017, which represents about 85.3 percent of its final three-year projections (Figure II.1). The original three-year enrollment target was 11,525 participants. However, the awardee revised this target down to 8,200 participants in the second program year because they felt that was more realistic given the pace of program enrollment during the first program year. When measured against the three-year projection set at the end of Year 1 (enough time for the awardee to gain a better understanding of the number of eligible participants but before the awardee began planning for the end of its cooperative agreement), the awardee met 60.7 percent of its projection.

Detroit Medical Center began enrolling participants in the Gateway program as soon as it was operational in January 2015. According to the enrollment data from the implementation and monitoring contractor and the awardee's own self-report data, program enrollment has steadily increased since January 2015.

Although Detroit Medical Center made good progress toward its revised three-year enrollment target, self-reported data suggests that they could have been more effective at enrolling participants. For example, Detroit Medical Center reported that its walkover rate—defined as the number of Gateway-eligible patients who presented in the ED as low acuity and were offered the opportunity to enroll in the program—fluctuated significantly over the three-year cooperative agreement. The awardees walkover rate was highest in the sixth program quarter (82.4 percent) and lowest in the tenth program quarter (22.5 percent). Detroit Medical Center believes that enrollment effectiveness could have been improved had this rate been more consistent over the life of the program.

Detroit Medical Center served participants with private insurance as well as participants who are Medicaid and Medicare beneficiaries. However, private insurance participants fall outside of the scope of our evaluation.

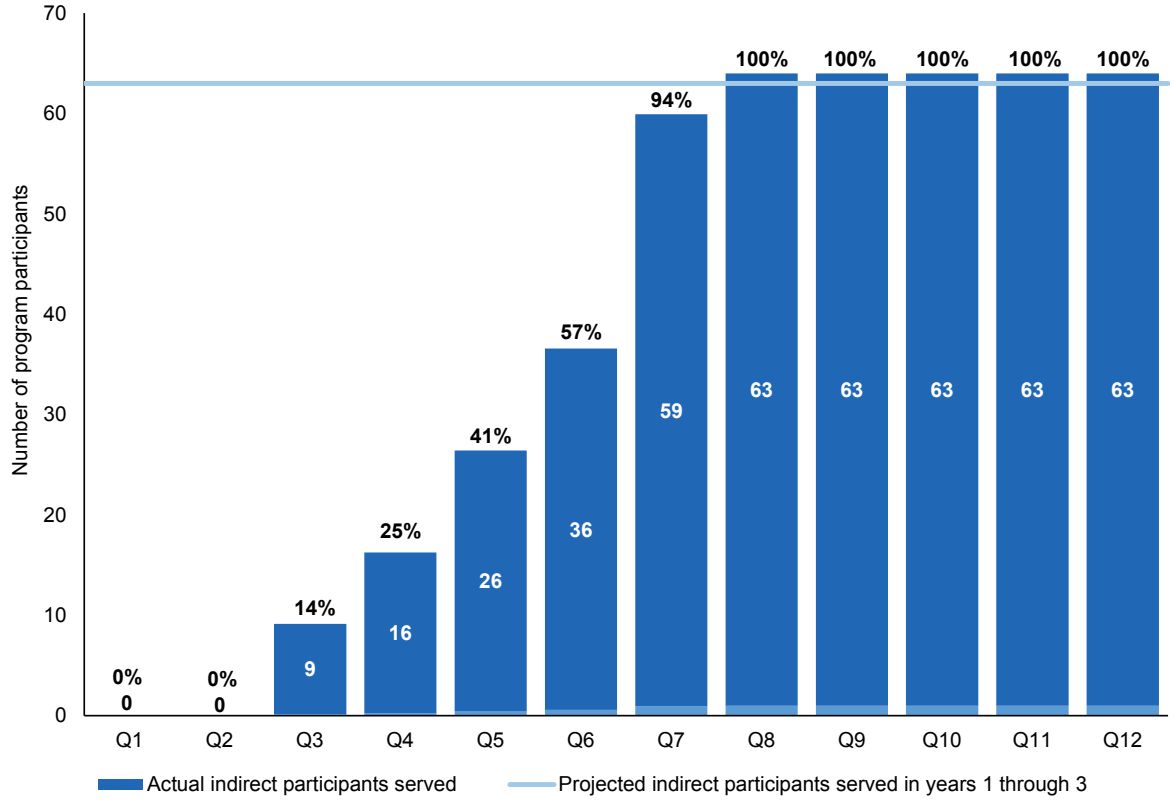
Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee revised their three-year enrollment target from 11,525 to 8,200 in the second year of the program.

Figure II.2. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers.

c. Barriers and facilitators associated with enrollment effectiveness

Detroit Medical Center encountered several barriers to meeting its three-year enrollment goal. For example, they experienced difficulty with gaining and maintaining buy-in from ED leaders and staff throughout the life of the Gateway program. Because Detroit Medical Center redirects patients from the ED to Gateway clinics for enrollment, the engagement and support of ED staff has been crucial to enrollment effectiveness. Detroit Medical Center made it a priority to develop strong working relationships with ED leaders and to identify champions within the ED to dispel the belief among ED staff that the Gateway program is taking patients away from the ED.

In addition, ED triage nurses played a critical role in participant enrollment because they were responsible for conducting the uniform triage assessment and MSE. The results of those assessments ultimately determine whether a patient should be connected to a Gateway navigator for the opportunity to enroll in the program. Detroit Medical Center works directly with the ED triage nurses to provide training on how to conduct the MSE. However, due to the demanding

nature of working in the ED, staff turnover is common. Thus, Gateway program leaders had to be diligent in training new ED triage nurses in performing the MSE.

Detroit Medical Center also noticed ED triage nurses were inconsistently administering the MSE at various times throughout the Gateway program. For example, in the third program year the number of MSEs resulting in a score of 4 or 5 (which indicates that a patient should be connected to a Gateway navigator) plummeted, whereas the number of MSEs resulting in a score of 3 (which indicates that a patient must remain in the ED) increased dramatically. Detroit Medical Center suspected that ED triage nurses might have been attempting to retain patients in the ED. To resolve the problem, Gateway program leaders placed an emphasis on strengthening relationships with the ED staff and reengaging ED champions in the third program year to promote program enrollment efforts.

A significant facilitator to enrollment effectiveness was the awardee's ability to use existing patient tracking systems to identify and recruit eligible patients into the Gateway program. The awardee uses FirstNet, which is an emergency information system that stores patients' electronic ED records and allows Detroit Medical Center staff to monitor patients on an interactive tracking board throughout their ED visit. Gateway program leaders enlisted the help of the Detroit Medical Center's information technology (IT) department to generate a list of patients in Detroit Medical Center's EMR system that met the Gateway program eligibility criteria and tag all patients who meet the eligibility criteria as "Potential Gateway" in FirstNet. Streamlining the identification of eligible patients in FirstNet helped jumpstart the enrollment process.

In addition, Detroit Medical Center expanded the Gateway program eligibility criteria in the second program year to accelerate progress toward its three-year enrollment target. As a result, populations such as nursing facility residents, foster youth, and Detroit Medical Center employees became eligible for enrollment in the Gateway program. Detroit Medical Center employees represent the population that resulted in the greatest boost to enrollment. However, there was limited enrollment uptake among the other two newly eligible populations of foster youth and nursing facility residents. Gateway program leaders suspected that this lack of uptake was due to the overwhelming complex medical and social needs often faced by individuals in these populations, making the Gateway program a lower priority for them.

2. Delivery of program services

a. Description of and changes to service delivery model

Detroit Medical Center implemented the Gateway program to reduce its ED utilization rates and to provide primary care services to patients with chronic health conditions. The Gateway clinics acted as PCMHs for Detroit Medical Center patients with chronic health conditions and high rates of ED utilization. Gateway program leaders modeled their PCMH after the Southcentral Foundation Nuka System of Care, a model of health care redesign.[4] The PCMH model provides services for targeted individuals' health care needs, replacing high-cost episodic care received in the ED with coordinated, long-term primary care. The Gateway PCMHs employ NPs or physicians who serve as primary care providers. Detroit Medical Center originally intended to only hire physicians to serve as primary care providers, however, the extended hours

[4] See https://www.southcentralfoundation.com/nuka/.

of operation and weekend availability proved to be undesirable to prospective hires. As a result, the awardee decided to hire NPs into the primary care provider role as well. In addition, the Gateway PCMH care teams include a registered nurse care manager, certified medical assistants, behaviorists, social workers, nutritionists, and pharmacist educators.

Detroit Medical Center originally operationalized four Gateway clinics: one at Sinai Grace Hospital (SGH), one at Children's Hospital of Michigan (CHM), and two at Detroit Receiving Hospital (DRH), all of which are hospitals operated by the Detroit Medical Center. One of the DRH Gateway clinics was originally housed within the Rosa Parks Geriatric Center, where elderly patients received care. However, the clinic lacked the space needed to effectively run the Gateway program parallel with the Rosa Parks clinic. Thus, at the end of the first program year, Detroit Medical Center decided to integrate the Rosa Parks Gateway Center into the main DRH Gateway Center, thereby allowing the site to run two shifts, offer longer hours of operation for patients, and usually treat geriatric patients early in the day. In addition, during the third program year, Detroit Medical Center made the decision to close the Gateway clinic at CHM. Gateway program leaders felt that the distinct needs of the pediatric population did not work well with the Gateway care model and thus this population was not a good fit for the program. As of the end of the third program year, only the SGH Gateway clinic and the consolidated DRH Gateway clinic remain operational.

Detroit Medical Center leaders believed that increasing Gateway participants' access to a usual source of primary care would lead to better management of chronic conditions and fewer ED visits. The awardee sought to improve the quality of primary care provided to this high-need, high-cost, underserved population. In doing so, it was expected that the Gateway centers would reduce health care costs not only by reducing the number of ED visits but also through care that was higher in quality; better coordinated; and patient-centered, such that it provided a broad range of clinical and social services.

> "Integral to [the] success [of the Gateway service delivery model] is that we have all these additional resources.… It's not just a physician and a medical assistant walking in and talking to the patient.… The fact that we keep our arms around the patient—with the care coordination component and the social work component and the mental health component—is why [the Gateway service delivery model] is so successful."
>
> —Gateway program leader

b. Evidence of service delivery effectiveness

Detroit Medical Center has been effective in delivering services through the Gateway program. This is evidenced by the awardee's implementation of the Gateway care model with very few major changes over the three-year cooperative agreement period, consistent staffing levels and high rates of staff satisfaction, and the ability to keep participants engaged throughout the life of the program. Although Detroit Medical Center experienced some setbacks in terms of closures of two of the original four Gateway clinics and periodic staff turnover, the awardee was able to deliver services through the Gateway program as planned. According to self-reported data, Gateway program leaders and staff are confident that the Gateway program has positively impacted health outcomes for Gateway program participants.

Delivery of intervention services. Detroit Medical Center has been effective in delivering intervention services. One indicator of this is Detroit Medical Center's self-reported data

regarding the availability of appointments at the Gateway clinics. Detroit Medical Center measured availability of appointments by calculating the percentage of Gateway participants who had their appointments rescheduled due to unavailability of Gateway providers. Based on data submitted by Detroit Medical Center in September 2017, the rate of rescheduled appointments due to provider unavailability was 0.2 percent.

In addition, Gateway program leaders reported that many participants have experienced improved chronic disease management and subsequent health outcomes in the time they have been enrolled in the program. For example, in the third program year Detroit Medical Center implemented a diabetes support group for adult Gateway participants that have uncontrolled diabetes. The support group is facilitated by a variety of Gateway providers, including NPs, dieticians, and social workers. The purpose of the support group is to provide disease management and medication education to Gateway participants and to introduce realistic lifestyle modifications to enable healthy eating and daily exercise. Participants are also encouraged to share their thoughts and concerns related to uncontrolled diabetes, as well as challenges they encounter with diabetes management. Participants that attend the support group are asked to complete a brief survey designed to measure their experience. As reported by Detroit Medical Center, participants have had a mostly positive response to the implementation of the support group. In addition, Detroit Medical Center monitored health outcomes for support group participants and found that they experienced significant improvement in their hemoglobin A1c and cholesterol levels, both of which are indicators of uncontrolled diabetes.

Staffing and training. Detroit Medical Center was successful in delivering Gateway program services partly due to high rates of staff satisfaction. For example, of the 20 Gateway program staff who responded to our HCIA R2 Staff Survey, 90 percent expressed satisfaction with their current role with the Gateway program. In addition, 75 percent of respondents reported that they had received formal training related to the Gateway program and 87 percent felt that the formal training helped improve their job performance. Gateway program leaders also reported overwhelmingly positive responses to their own staff surveys in both the second and third program years.

> "[The Gateway diabetes] support group entails providing patient education to some of our highest-risk diabetes patients in a group setting…. [The support group sessions] were led by a physician, nurse case manager, and pharmacist educator. The first [session] was very valuable. We had six people attend … and they really enjoyed it. They started to talk and help one another…. The [support] group sessions oftentimes help some [patients] in a different manner. It's not one-on-one [with a clinician]. It may help some patients more when [a peer] says it."
>
> —Gateway program leader

Infrequent staff turnover also played a role in Detroit Medical Center's success in delivering Gateway program services. Gateway program leaders reported that staffing has been consistent throughout the three-year cooperative agreement, with only a handful of Gateway providers departing the program. In addition, Detroit Medical Center was able to fill most staff vacancies in a timely manner.

Engagement of providers. Detroit Medical Center's engagement of the Gateway program's primary care providers influenced the awardee's success in delivering Gateway program services. There was very little turnover amongst its primary care providers over the three-year cooperative agreement and there were high levels of provider satisfaction. To engage the Gateway primary care providers, the Gateway program's principal investigator and two medical directors conducted a utilization review process with the

Gateway providers. The reviews served to continue the clinical education of the providers by having them participate in case studies and learn how to handle complex medical cases, thereby enhancing their decision-making capabilities. In addition, in the interviews we conducted with non-clinical Gateway program staff, the primary care providers were described as crucial to the success of the program because the providers played a key role in leading the care teams at the Gateway clinics by providing necessary leadership, clinical expertise, and general support to the rest of the Gateway program staff.

Engagement of program participants. Detroit Medical Center's engagement of Gateway participants also contributed to the awardee's success in delivering Gateway program services. Gateway program leaders identified the appointment no-show rate as a critical issue for participant engagement early in the program. The awardee implemented various strategies to improve the no-show rate, including phone reminders and follow-up calls after missed appointments. According to the Detroit Medical Center's self-monitoring data, the no-show rate was highest in the second program quarter at 72.3 percent. The no-show rate has steadily decreased since then and as of the 12th program quarter was 59.5 percent.

Participant engagement was also evidenced by the number of return visits at the Gateway clinics. Detroit Medical Center reported that as of the 12th program quarter, 61.5 percent of all visits to Gateway clinics were follow-up visits.

c. Barriers and facilitators associated with service delivery effectiveness

The awardee Center experienced several barriers related to service delivery effectiveness over the three-year cooperative period, the most significant of which were the challenges that arose from the change in ownership of Detroit Medical Center. Tenet Health acquired Detroit Medical Center from its previous owner, Vanguard Health Systems, in late 2013. In the years that followed the acquisition, Detroit Medical Center transitioned its information systems provider to match that of its parent company, Tenet Health. As a result, the awardee experienced challenges with collecting self-monitoring data. Under Vanguard Health Systems, Detroit Medical Center used an in-house information systems provider; an external provider is now used under Tenet Health. Gateway program leaders reported that an advantage of using an in-house provider was that reports were highly customizable. After the transition, program leaders were not able to obtain customized reports from the new provider. In addition, Gateway program leaders reported that they have experienced significant delays in their requests for data under this new provider.

The lack of available space also presented a challenge to service delivery effectiveness. The Gateway clinics are colocated with or adjacent to busy EDs and thus available space for expansion of the clinics was limited. As participant enrollment increased steadily over the three-year cooperative agreement, Detroit Medical Center was unable to scale up the amount of physical space to accommodate the growing operations at the Gateway clinics. According to interviews with Gateway program leaders, the lack of available space at times resulted in long appointment wait times—sometimes approaching 60 minutes—due to the occupation of all available patient rooms.

The most significant facilitators associated with service delivery effectiveness of the Gateway program were the stability in staffing levels and high rates of staff satisfaction over the

three-year cooperative agreement. Detroit Medical Center experienced some staff turnover, however, Gateway program leaders very rarely struggled to fill vacant positions. In the third program year, Detroit Medical Center terminated a NP at the DRH Gateway clinic due to poor job performance. Gateway program leaders experienced a delay in replacing the NP at the DRH Gateway clinic, which resulted in a temporary reduction in patient volume at the clinic. As a stopgap, Detroit Medical Center contracted with one of the Gateway medical directors to serve as a part-time Gateway provider until a replacement NP was hired. Once Detroit Medical Center brought on a replacement NP, patient volume returned to previous levels.

Gateway program staff identified several factors that influenced their positive perceptions of the program. The most commonly cited factors that influenced the Gateway program's ability to achieve its goals included participant engagement in the program, buy-in to the program by staff, and a shared understanding of program goals. The ability of program team members to operate efficiently and collaborate effectively to meet participants' needs were commonly cited by respondents as facilitators to the Gateway program achieving its goals.

Another facilitator associated with service delivery effectiveness was Detroit Medical Center's ability to collect self-monitoring data. Gateway program leaders benefited from the existing information systems infrastructure in place at Detroit Medical Center that was used to collect data on the Gateway program. Although the awardee experienced aforementioned challenges associated with transitioning information systems providers during the cooperative agreement period, the awardee still benefitted greatly from the data reports that were generated though various information systems platforms. The primary information system Gateway program leaders relied on for data collection was the Cerner system, which is also the EMR system that Detroit Medical Center currently uses. Gateway program leaders use Power Insider Business Objects, an operational analytics software tool, to extract information from the Cerner system regarding Gateway appointment information. For example, program leaders were able to use this software to determine the number of appointments made by Gateway participants, the number of no-show appointments, and the average length of time of Gateway appointments.

Gateway program leaders have also been able to access financial management data collected by the Detroit Medical Center finance department. This data includes billing and reimbursement information. Gateway program leaders have used this information to determine whether the Gateway intervention has resulted in a reduction in the cost of providing care to Gateway participants.

Finally, Detroit Medical Center used Healthy Maintenance, which is a population health management application that integrates with the Cerner EMR system. Gateway program leaders have used this application to track health outcomes for Gateway participants based on demographics, chronic conditions, and other patient characteristics. Based on the self-monitoring data from the Healthy Maintenance tool, Detroit Medical Center made the decision to implement a support group for diabetic patients after noticing that patient health outcomes for these Gateway participants with diabetes could be improved.

C. Assessment of perceived program effects on the delivery of care and outcomes

The Gateway program staff who responded to our survey felt that the Gateway program had positive impacts on its participants. All of the respondents reported that they felt the program had been effective in achieving its goal of reducing ED overutilization and improving the disease management of Gateway participants. The majority of respondents felt that the Gateway program was making a difference in meeting the critical needs of patients in the Detroit metropolitan area. Specifically, staff felt that the Gateway program had a positive impact on program participants by improving their access to care, coordination of health care services, quality of life, and health outcomes.

Detroit Medical Center's self-reported Gateway participant ED utilization data also suggests that Detroit Medical Center has been effective in delivering Gateway program services. Detroit Medical Center reported a reduction in the frequency with which Gateway participants were presenting in the ED. The awardee believed this was due to the fact that the Gateway program provided participants with access to primary care and follow-up care services.

> "The individual impact is huge…. We hear the most amazing stories from patients about their experiences.… Our look at the data would suggest that we're seeing reduced frequency in our patients of utilizing the ED for routine care or primary care…. I think that a big impact has really been … how Gateway has established a name in the community and across [Detroit Medical Center] as not being a threat but being something that is embraced and is being recognized as providing very important primary care services to the community."
>
> —Gateway provider

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. In our interviews, Gateway program leaders, primary care providers, and non-clinical staff all reported that they were confident that the Gateway program positively impacted participants. They felt that through coordinated ongoing primary care services, the Gateway program improved the chronic disease management of participants. Gateway program leaders felt that the improved disease management will serve to reduce participants' need to use the ED. The awardee also reported that, in some cases, participants' health outcomes appeared to improve from the time they first enrolled in the Gateway program.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Detroit Medical Center's Gateway program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Detroit Medical Center received a three-month no-cost extension and will continue to operate until December 31, 2017. The projected sample sizes in Table III.1 are based upon actual enrollment as of August 31, 2017, of which most but not all participants will receive six months of program exposure, a requirement for inclusion in our evaluation to minimize dilution of program effects due to limited exposure to the program components. Because of processing lags in Medicaid data, we have not confirmed that all 2,365 Medicaid beneficiaries meet program eligibility for inclusion in any analysis.

Table III.1. Assessment of HCIA R2 awardee evaluability as of August 31, 2017: Detroit Medical Center

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure	432[a]
Projected Medicaid population with 6 months of program exposure	2,365[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	2,563
Likelihood of all-cause hospitalizations	1,423
MDE sample size requirement to detect 20% effect	
Total expenditures	641
Likelihood of all-cause hospitalizations	356
Participation/selection bias of concern	Yes, patient self-selection high/high refusal rate
Intervention implemented that differs from baseline period	Questionable, cannot identify degree/intensity of intervention
Claims sufficient to identify treatment and comparable comparison group?	Questionable, no testing yet to determine strength of intent-to-treat framework
Likelihood of solid comparison group	Serious concern. We may not be able to identify a strong comparison group.

Table III.1 *(continued)*

Evaluability domain	Response
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Lack of adequate comparison group
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in any analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

At this point, we do not expect to conduct a rigorous impact evaluation of the Gateway program. The primary concern is constructing a solid comparison group, given the very low rate of participation among those who are flagged as meeting the initial eligibility criteria: frequent ED use, no usual source of primary care, and consent to participate. To date, the awardee has identified almost 37,500 potential participants based upon a review of its EMR for frequent use of its EDs. However, less than 15 percent of these participants have enrolled in the program because (1) they were not recruited before they left the ED and could not be contacted after discharge; (2) they were found to be ineligible because they stated that they had a primary care provider, who was not identified in the awardee's EMR; or (3) they refused to enroll in the program. We have serious reservations about our ability to develop a claims-based algorithm that identifies such a small percentage of potential participants within an intent-to-treat framework because it would lead to significant dilution of intervention effects. We will attempt to devise a claims data algorithm—using observations of eligible and ineligible Detroit Medical Center patients—to assess the likelihood that participants were contacted, had a primary care provider, and agreed to participate. Given the small number of Medicare FFS beneficiaries, we will consider this approach for Medicaid-only beneficiaries and report our results in the final report if T-MSIS data become available and we are confident that we can produce unbiased program estimates. At this point, we do not have any awardee-specific data on implementation to report. We will report on the experiences of staff and participants, based on our surveys.

B. Characteristics of Medicare and Medicaid participants at baseline

For the purpose of our evaluation, the treatment group consists of beneficiaries recruited for the Gateway program who either (1) used the Detroit Medical Center's EDs for medical care at least five times during the past year or (2) had at least one of seven chronic conditions and were referred to the Gateway program without having had an ED visit. Beneficiaries who indicated that they had an ongoing relationship with a primary care provider were deemed ineligible for the program and were referred back to that provider. The treatment group also includes patients who enrolled in Gateway as well as recruited patients who met the eligibility criteria but refused to enroll.

Because we are conducting an intent-to-treat evaluation, we are defining the date of the initial recruiting attempt as the participant's enrollment date, although for some who accepted the invitation to enroll in Gateway, the date of the first visit to a Gateway clinic or the formal signing of the institutional review board form occurred later. For patients who were referred to Gateway, the date of their first visit is the beginning of their treatment period. For those whose recruitment was triggered by an ED visit, any costs associated with that visit (or with a subsequent

hospitalization) were assigned to the baseline year, and the treatment period began on the day after their ED visit or after they were discharged from the hospital.

In presenting the baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they must have been recruited for the awardee's program on or before May 31, 2016. The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. After we excluded beneficiaries who did not meet the above criteria, 162 participants were included in the analysis of baseline characteristics for this report, 93 of whom (57 percent) were successfully recruited and enrolled.[5]

The Medicare FFS beneficiaries recruited for the Gateway program are a predominantly disadvantaged and high-needs group (Table III.2). A sizable majority of them are under the age of 65 (61 percent); only 1 percent are over the age of 85. In total, 7 in 10 beneficiaries were originally eligible for Medicare due to a disability. Nationally, only 24 percent of beneficiaries became eligible due to a disability. Thus, the awardee is recruiting a population that generally has significant health care needs and high Medicare expenditures. Consistent with this, over half (51 percent) of the recruited beneficiaries are dual eligibles, which indicates a high level of social need. This compares with 15 percent in the national Medicare FFS population.

Nearly all recruited beneficiaries are African American (93 percent), reflecting the racial composition of Detroit. In comparison, only 11 percent of beneficiaries nationally are African American. The average hierarchical condition categories (HCC) risk score (1.5) is 50 percent higher than it is among all Medicare FFS beneficiaries, suggesting that Gateway participants are in substantially poorer health and have greater needs for care than does the general Medicare FFS population.

Consistent with these high-need characteristics, treatment group beneficiaries had high rates of service use and Medicare expenditures in the year before enrollment. In Table III.3, we report baseline utilization and expenditure data for a common set of measures, including the four core measures from CMMI. By offering convenient, patient-centered primary care services in the Gateway clinics, the Detroit Medical Center expects to reduce the utilization of high-cost services in EDs and hospital settings compared to what would have occurred absent the program. If reductions in the utilization of these high-cost services are realized, then there would be a corresponding decline in Medicare expenditures. We examined the baseline costliness of care by

[5] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

calculating average per beneficiary per month (PBPM) Medicare payments—in total and by major types of services. The average total PBPM Medicare payment during the baseline year was $1,503. By comparison, the national average total PBPM expenditure in 2013 ($792) was about half of the awardee's PBPM average expenditure. Average PBPM Medicare payments for acute inpatient care ($614) and physician services ($382) were the largest drivers of the total cost of care. Outpatient care, which includes facility charges for ED visits not resulting in a hospital and/or an observation stay, contributed $226 to the total PBPM cost.

About one-third of the recruited beneficiaries in the treatment group (34 percent) were hospitalized during the baseline year. The overall annual rate of acute care hospitalizations was 841 per 1,000 Medicare FFS beneficiaries during the baseline year, with a 41 percent likelihood of a 30-day readmission per hospital discharge among those hospitalized. The readmission rate is more than twice the national average of 18 percent for Medicare FFS beneficiaries, as reported by CMS. We would expect nearly all recruited beneficiaries to have had an ED visit, and 79 percent had at least one visit that did not result in a hospitalization. Beneficiaries in the treatment group averaged 4.3 ED visits during the baseline year. The corresponding rate for outpatient ED visits that did not result in a hospitalization was 3.2, suggesting that these beneficiaries often visited the ED when they had low-acuity conditions, many of which may have been treatable in physicians' offices.

At baseline, the annualized rate of primary care visits (6,641 per 1,000 Medicare FFS beneficiaries) was substantially lower than the rate of specialty visits (13,081 per 1,000 Medicare FFS beneficiaries), suggesting a possible need for greater access to primary care, which is a primary focus of the Gateway program. More than half of primary care and specialist visits (56 percent and 59 percent, respectively) take place in ambulatory care settings. As would be expected given the fact that recruitment occurred in EDs, there was dramatically higher inpatient and ED use during the fourth quarter of the baseline year. For instance, the rate of hospitalizations per 1,000 beneficiaries rose from 842 in the earliest quarter of the baseline year to 1,313 in the quarter that ended in Gateway recruitment. The corresponding rates for outpatient ED visits were 3,250 and 5,096. This suggests that the beneficiaries' health conditions are becoming worse, thus increasing the likelihood of ED visits and, in turn, opportunities for recruitment into the Gateway program.

Finally, we calculated the prevalence of the seven chronic conditions used in the eligibility criteria for the program (see Table III.4). The conditions are based on Chronic Conditions Warehouse (CCW) definitions. Not surprisingly, the treatment group had higher chronic condition rates than those in the full Medicare population. For instance, the prevalence of asthma was nearly 3 times greater in the treatment group than in the general Medicare FFS population as a whole (14 percent versus 5 percent). Over half of the beneficiaries in the treatment group had hypertension (54 percent), slightly below the national average. Medicare beneficiaries in the treatment group averaged 1.6 of these seven chronic conditions, with 31 percent not having any of the seven conditions in 2014.

We also compared chronic condition prevalence rates for beneficiaries who accepted recruitment into the Gateway program with the rates for those who did not. Although the prevalence rates of several conditions varied from 5 to 10 percentage points, neither participants nor refusers consistently had higher prevalence rates. It is likely that this inconsistent pattern is

an artifact of the relatively small size of the two subsamples. The overall burden of chronic conditions on patients was similar between the two groups.

Although the HCC scores of participants and refusers were similar, the average PBPM Medicare payment for participants was only $1,283 compared with $1,811 for those who declined recruitment. These cost differences are reflected in higher inpatient, hospital outpatient, and physician costs. The rate of readmissions for hospitalized refusers was twice as high as the rate for participants during the baseline year (52 versus 24 percent). It will be important to see whether these patterns persist after the treatment group size grows larger.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Detroit Medical Center's program through May 31, 2016

	All participants (N = 162)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	99	61
65 to 74	47	29
75 to 84	14	9
85 and older	2	1
Gender		
Female	72	44
Male	90	56
Race		
White	11	7
Black	151	93
American Indian, Alaska Native, Asian/Pacific Island American, or other	Not applicable	Not applicable
Hispanic	Not applicable	Not applicable
Original reason for Medicare eligibility		
Old age and survivor's insurance	45	28
Disability insurance benefits	115	71
End-stage renal disease (ESRD)[a]	2	1
Hospice[b]		
Medicare/Medicaid dual status, percent dual[b]	83	51
HCC score[c]		Statistic
Mean		1.5
25th percentile		0.56
Median		1.09
75th percentile		1.71

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Table III.2 *(continued)*

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Detroit Medical Center's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	162	143	154	159	162
Average Medicare expenditures PBPM[a]					
Total	1,503 (221)	1,075 (305)	1,196 (241)	1,457 (312)	2,221 (413)
Acute inpatient	614 (125)	383 (224)	529 (179)	478 (213)	1,034 (239)
Inpatient other[b]	127 (50)	74 (45)	0 (0)	212 (96)	211 (125)
Outpatient[c]	226 (42)	220 (61)	176 (37)	201 (46)	303 (57)
Physician services	382 (43)	307 (44)	351 (47)	402 (56)	460 (58)
Home health	101 (20)	81 (32)	87 (22)	104 (28)	129 (27)
Skilled nursing facility	41 (16)	6 (27)	25 (25)	55 (35)	73 (37)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	12 (5)	4 (2)	29 (20)	5 (2)	10 (4)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	841 (174)	578 (248)	701 (232)	735 (347)	1,313 (253)
Outpatient ED visits	3,250 (488)	2,974 (1,272)	2,346 (491)	2,495 (507)	5,096 (589)
Total ED visits	4,320 (699)	3,552 (1,254)	3,101 (674)	3,598 (800)	6,872 (805)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Observation stays	431 (163)	289 (186)	297 (120)	552 (259)	566 (161)
Primary care visits in any setting	6,641 (841)	4,996 (957)	5,232 (776)	7,879 (1,298)	8,236 (1,139)
Primary care visits in ambulatory settings	3,728 (397)	3,379 (513)	3,344 (494)	4,307 (514)	3,835 (534)
Specialist visits in any setting	13,081 (1,665)	9,905 (1,892)	11,029 (1,666)	13,105 (2,303)	17,836 (2,424)
Specialist visits in ambulatory settings	7,684 (717)	6,006 (1,043)	6,742 (898)	6,776 (871)	10,964 (951)
Measures of any health care utilization					
Percentage with a hospital admission[d]	34 (4)	8 (2)	9 (2)	12 (3)	23 (3)
Percentage with an outpatient ED visit[e]	79 (3)	26 (4)	26 (3)	31 (4)	64 (4)
Percentage with an observation stay[f]	16 (3)	2 (1)	5 (2)	6 (2)	11 (2)
Percentage with a 30-day readmission among all discharges	41 (5)	36 (12)	39 (12)	52 (9)	28 (8)
Percentage of participants with a readmission among all participants	6 (2)	1 (1)	2 (1)	3 (1)	3 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

Table III.4. Prevalence of chronic conditions in 2014 among Medicare FFS beneficiaries in Detroit Medical Center's treatment group, by whether they accepted recruitment (percentages unless otherwise indicated)

Chronic conditions	Total treatment group	Accepted recruitment into program		National Medicare prevalence
		Yes	No	
Asthma	14	15	13	5
Heart failure (CHF)	15	14	16	14
COPD	19	22	15	11
Diabetes	28	25	32	28
Hypertension	54	53	55	57
Depression	26	29	23	17
HIV/AIDS	4	6	2	0.4
Number of observations	**149**	**87**	**62**	

Source: Mathematica analysis of information from the awardee's finder file and Medicare claims and enrollment data as of May 31, 2016. National prevalence rates are from https://www.ccwdata.org/web/guest/medicare-charts/medicare-chronic-condition-charts except for HIV/AIDS rates, which are from the Kaiser Family Foundation Fact Sheet, "Medicare and HIV," October 2016 (http://files.kff.org/attachment/Fact-Sheet-Medicare-and-HIV).

Notes: The seven chronic conditions are targeted by the Gateway program and serve as eligibility criteria. The prevalence is based on 2014 CCW chronic "ever had" condition flags and may not precisely replicate the clinical definitions used by DMC to identify patients with specific chronic conditions. The sample hence excludes any beneficiaries in the treatment group who were not Medicare FFS beneficiaries for all 12 months of 2014. DMC added two conditions after the start of the program, but they apply only to children.

CCW = Chronic Conditions Warehouse; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Detroit Medical Center planned to support its program through existing FFS billing and did not propose a new payment model.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: (1) the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor and (2) narratives and payment model reports the awardee submitted to the implementation and monitoring contractor. We did not interview the awardee about its payment model because it did not advance from the development stage to active negotiations with one or more payers.

B. Description of the payment model

Detroit Medical Center did not submit a proposal for a new payment model and planned to continue to support the program through billing traditional FFS and transitional care management codes. The awardee noted that the services provided under the program are reimbursable under current primary care payment structures and did not require a new model.

C. Status of the payment model

The awardee explored shared savings, enhanced FFS, and hybrid payment models with several payers with Medicare, Medicaid, and commercial lines of business, but did not submit a proposal for a specific value-based model as of May 2017, the end of eleventh program quarter of the three-year cooperative agreement. The awardee planned to continue to meet with payers and physician-hospital organizations that act as accountable or quality improvement partnerships to explore opportunities for the program to contribute to their goals and be incorporated into value-based payments.

D. Factors associated with the development of the payment model

Detroit Medical Center was able to engage with several payers who were interested in the program's goal of reducing avoidable ED visits. However, the awardee was not able to advance the discussions to an agreement on a value-based model.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Detroit Medical Center expected to sustain its Gateway program at two of its three original sites after the cooperative agreement ended, through internal support and FFS payments. Detroit Medical Center's parent, Tenet Health, agreed to provide support funding to continue the program at two of the three original sites for another year. The awardee added a component and was streamlining others as a strategy to gain support from its parent and potentially other funders. The awardee was working to secure relationships with other providers—both primary care and hospitals—to sustain and potentially scale the program throughout the city, but struggled to demonstrate how the program directly helps other hospitals. The program had not been replicated elsewhere, although the awardee was starting to think about how it could foster replicability.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, Detroit Medical Center had begun actively pursuing strategies to sustain the Gateway program. The awardee was holding discussions with multiple partners—commercial payers, the state Medicaid program, Medicare, and local accountable care organizations—about

developing a feasible payment model. The program had not been scaled or replicated, but Detroit Medical Center conducted an outreach campaign to promote replication of the Gateway program. To inform other providers about the program, the awardee sent mailings to approximately 175 PCPs in the Detroit area and followed up with them by phone. They offered training modules to these providers, but received very little interest. Most providers reportedly said that they did not want to implement a new care model that did not offer clear financial incentives.

C. Implementing the SSR plan: progress and changes

Sustainability. In Year 3, Detroit Medical Center committed to continuing the program's current staffing levels at two of the three original sites for another year (after the end of its no-cost extension in December), in part because Tenet Health (its parent company) agreed to provide resources to support the program through December 2018. As discussed in the Payment Model chapter, Detroit Medical Center did not propose a new payment model and instead planned to continue to support the program by billing traditional FFS and transitional care management codes. The awardee tried to secure support from payers for an advanced payment model, but those discussions stalled. The awardee reported wanting to form more collaborations with payers that participated in both Medicaid and commercial business, presumably as a way to generate more revenues.

The awardee did not expect the program to change significantly after the award, but wanted to improve its efficiency. It would maintain the patient navigator and transitional care elements, but track patients differently. The awardee also planned to streamline program processes. For example, the awardee planned to "optimize every resource" and improve the patient flow process so that patients would not wait too long. This would potentially mean involving physicians in different ways, such as having physicians on site in the clinics to manage the patient load, staff and operational issues (one site had implemented this model, while the other site had an NP clinical lead). However, some program stakeholders thought the program should change its structure going forward. For example, one respondent suggested maintaining the primary care piece of the program as is, but incorporating new urgent care strategies to more readily address the situations and gaps in care and situations that might trigger hospital use.

To obtain Tenet's "total buy-in," the awardee reported that they would need to demonstrate that the program results in cost avoidance for the provider by reducing ED overutilization and hospital readmissions. As one respondent said, "It's one of those things they [Tenet] might not notice what we do for them right now, but if we were to go away, they would feel it really quickly." Adding the transitional care component was a key strategy to help demonstrate to Tenet that the program generates financial benefits to the system. Detroit Medical Center incorporated this component at one Gateway clinic and planned to incorporate it at the other clinic. In addition, Detroit Medical Center was also working to improve the transitional care component by strengthening relationships with other community providers and organizations. This proved challenging because many had similar initiatives.

Scalability. Detroit Medical Center was collaborating with Chrysler Fiat UAW, which was interested in establishing better access to primary care for employees who were relying on the ED for primary care. The awardee had given multiple presentations to the employer and

employees about replicating the Gateway program. Ultimately, Chrysler decided to contract directly with Detroit Medical Center to provide the services.

Replicability. Although the Gateway program had not yet been replicated, there were activities that signaled the awardee's interest in helping others adopt the program. As noted, the awardee's efforts in Year 2 to generate interest among other PCPs were not very fruitful. In Year 3, Detroit Medical Center's CEO spoke with entities in Chicago who are interested in replicating a program like Gateway. Also, the program planned to continue collecting patient consent for institutional review board purposes to allow for potentially publishing its findings.

D. Factors associated with progress toward implementing the SSR plan

The awardee reported that Tenet's willingness to support the program for a year was related to the program's reputation as being more efficient, providing more support services, and helping patients transition to appropriate follow up care, than other primary care areas of the system. "The fact that we keep our arms around the patient with the care coordination component and the social work and the mental health component of it is why, I think, we're so successful," said a Gateway program leader.

Another stated key factor to SSR was that because the program is effectively integrated into Detroit Medical Center's overall operations, others did not view it as a separate program. This presented pros and cons. The strong connections to the broader organization helped the awardee build internal support for the program. Also, some external entities were willing to work directly with Detroit Medical Center, viewing the program as another Detroit Medical Center incentive program linked to broader Detroit Medical Center goals to address over-utilization of hospital services. However, this also meant that gaining payment or other resources from other providers and payers directly for the program was difficult, and the awardee was concerned that resources provided to Detroit Medical Center broadly might ultimately not benefit the program.

Another challenge for the awardee was obtaining funds from payers and providers for a program that was currently funded. The awardee found that once a program provides a service for free (because the federal government had been paying for it), it was difficult to then go back and ask for funds to keep it going, especially without strong evidence of the program's financial benefits.

Detroit Medical Center tried to secure support from payers for an advanced payment model, but those discussions stalled. The awardee hired a consultant with significant connections with state officials and a strong data background, but initial meetings with payers were not fruitful. The awardee was working with the consultant to better package its messaging; integrate its data; and initiate new conversations with potential funders, especially the state Medicaid program because about half of its participants were Medicaid beneficiaries.

Also, the awardee thought that improving efficiency and developing relationships with other physician organizations would be key to long-term sustainability as payment types change (that is, the move to value-based payment).

A key challenge to scaling the program throughout Detroit was that fewer such potential practices remain in the city than the awardee's initial assessment indicated (about 60 compared

to 175). Many physicians had passed away or retired and new physicians had not taken over their practices, which the respondent attributes to the poor payer mix of the inner city and younger doctors not wanting to practice there. Many remaining practices did not have the resources to provide the functions (for example, open access scheduling) required for PCMH status.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension that allows it to continue to enroll new beneficiaries. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted above, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: New York City Department of Health and Mental Hygiene

April 18, 2018

Kyle Emery (RTI International)
Noah Triplett (RTI International)
Cordon Newhart (RTI International)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] DOHMH received a 6 month no cost extension. Enrollment ended on February 28, 2017 and services ended on August 31, 2017.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

The New York City Department of Health and Mental Hygiene (DOHMH) estimates that about 2.4 percent, or 146,500, New York City adult residents have the hepatitis C virus (HCV) and many of them are not receiving treatment.[4] To address this problem, The Fund for Public Health in New York and DOHMH implemented Project INSPIRE to improve HCV treatment by using patient-centered medical and behavioral health care integrated with care coordination and tele-mentoring. The Fund for Public Health in New York received HCIA R2 funding to create and support the implementation of Project INSPIRE. The Fund for Public Health in New York is a nonprofit organization established by DOHMH to advance the health of all New York City residents. Program leaders at DOHMH, the agency responsible for program implementation, anticipated that evidence of the program's effectiveness would support reimbursement for participants' services as well as the expansion of this model of care citywide.

DOHMH partnered with two health systems, Mount Sinai Health System and the Montefiore Health System, to implement Project INSPIRE. The project's care coordinators and peer navigators help participants to complete HCV treatment by addressing the underlying health problems that commonly interfere with adherence to treatment—including mental health and substance abuse issues—while teaching participants the skills they need to manage their health independently. Project INSPIRE also provided tele-mentoring for primary care and other providers to increase their capacity to treat HCV. Both health systems facilitated weekly training and the exchange of information among hepatologists and program providers through tele-medicine, a process DOHMH referred to as tele-mentoring. See Table I.1 for more information on the overall program.

DOHMH planned to serve 3,200 participants with a detectable HCV RNA viral load (in other words, having an active HCV infection) by the end of the three-year cooperative agreement. The awardee anticipated that Project INSPIRE would result in a sustained viral response (SVR), or HCV cure, among 90 percent of non-cirrhotic participants and among 50 percent of cirrhotic participants; DOHMH also hoped to achieve satisfaction among 80 percent of participants and to reduce the number of new episodes of acute care. Project INSPIRE's expected outcomes were to reduce Medicare and Medicaid costs, as indicated by total expenditures, hospitalizations, readmissions, and ED visits by doing the following: (1) preventing HCV infection from advancing to hepatocellular carcinoma or other forms of liver disease; (2) stabilizing and managing participants during the program; (3) improving participants' self-sufficiency by facilitating the treatment and management of comorbid conditions, such as HIV, substance abuse disorders, mental health disorders, diabetes, and heart disease; and (4) preventing repeated treatment for HCV by avoiding treatment failures and re-infections.

DOHMH implemented its program as planned, with very few changes to the enrollment process or delivery of services. The only significant change the awardee made was to revise the health promotion materials to make them more accessible to the target population, in which

[4] See http://www.hivforum.org/storage/documents/2015/Summit/Presentations/C1_05_Laraque.pdf.

many people have low levels of health literacy, according to feedback from care coordinators and peer navigators.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	To facilitate HCV treatment for participants by improving clinical and non-clinical care for both HCV and comorbid conditions and by using tele-mentoring to increase the capacity of health care providers to effectively treat HCV
Major innovation	Using care coordination to guide an underserved population through HCV treatment by addressing the underlying health problems that commonly interfere with adherence to treatment and by using tele-mentoring to enable primary care providers to treat HCV
Program components	• Care coordination • Tele-mentoring
Target population	HCV-positive individuals with a detectable HCV RNA viral load who were born between 1945 and 1965; who reside in the Bronx or in East or Central Harlem in New York City; who are eligible for Medicare or Medicaid; and who have difficulty keeping appointments, have received sporadic care, have never been in care, or who have requested support. Other adults age 18 and older who live in the five boroughs of New York City may also participate, although they are not being actively recruited.
Theory of change/ theory of action	• Improve HCV cure rates by increasing access to treatment, offering care coordination and health promotion services, and addressing comorbid conditions that can interfere with HCV treatment • Improve participant satisfaction by using a patient-centered service model • Decrease expenses by reducing the need for costly health interventions associated with complications of HCV or comorbid conditions
Payment model	Value-based payments, shared savings, bundled or episode payment, capitated payment for care management/coordination services
Award amount	$9,948,459
Effective launch date	January 15, 2015
Program setting	Health systems (Mount Sinai and Montefiore)
Market area	Urban
Market location	New York City
Target outcomes	• SVR or HCV cure rate • Participant satisfaction • Episodes of acute care for behavioral conditions

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. DOHMH successfully implemented Project INSPIRE's facilitation of HCV treatment by (1) coordinating services for patients with HCV who frequently fall out of care and (2) tele-mentoring providers who do not specialize in treating the liver on HCV treatment, allowing these providers to also prescribe HCV medication to their patients. We based this conclusion on five factors. First, the awardee enrolled 2,775 participants—95.2

percent of its enrollment target—by the end of the initial cooperative agreement. Second, the awardee was fully staffed and experienced little turnover. Despite sometimes large caseloads, the awardee achieved high rates of enrollment and treatment initiation. Third, participating providers at implementing sites were effectively engaged in the tele-mentoring component and worked with care coordinators to refer participants. Fourth, program participants were successfully engaged by care coordinators and peer navigators, which led to high rates of successful initiation and completion of the program. Last, participating clinicians and other implementation staff reported that the program had a positive effect on the delivery of care.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of DOHMH's Project INSPIRE, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. DOHMH and the two partnering managed care organizations (MCOs), HealthFirst and Visiting Nurse Service of New York (VNSNY), drafted a payment model consisting of a one-time bundled payment from Medicare or from Medicaid MCOs, funding care coordination services for patients with HCV. The two partnering MCOs are negotiating with Montefiore and Mount Sinai to implement the payment model. The awardee is also conducting a return-on-investment (ROI) analysis, which may help these negotiations.

Sustainability plans. DOHMH had made considerable progress on planning for sustainability. While awaiting full acceptance and implementation of its payment model, the awardee and its hospital partners gained a commitment and resources from hospital leadership to continue the program in the short term and were trying to obtain additional funding. At the same time, they modified the program slightly to contain its costs in some ways while making it more robust in other ways. The program had not been scaled or replicated although the awardee and its partners thought that it could be and pursued activities toward that end.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of non-clinician staff, including care coordinators and peer navigators, on their perceptions of the effect of the program on the delivery of care. This survey was fielded from July to October of 2016 and achieved a response rate of 77 percent with 31 respondents. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; 65 to 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

As mentioned, DOHMH partnered with two New York City health systems—Montefiore Health System and Mount Sinai Health System—to implement Project INSPIRE at 24 clinical sites. The awardee chose these two systems because of their expertise in HCV care, their ability to provide holistic care for program participants, and their location in the Bronx or in East or Central Harlem (areas with high rates of HCV).

To have been eligible for the program, a person must (1) be at least 18 years old; (2) live in New York City (the Bronx, Kings, New York, Queens, and Richmond counties); (3) be eligible for Medicare or Medicaid; (4) have a detectable HCV RNA viral load (in other words, have active HCV infection); (5) consent to participate in the program; and (6) have difficulty keeping appointments, have received sporadic care, have never been in care, or have requested support.

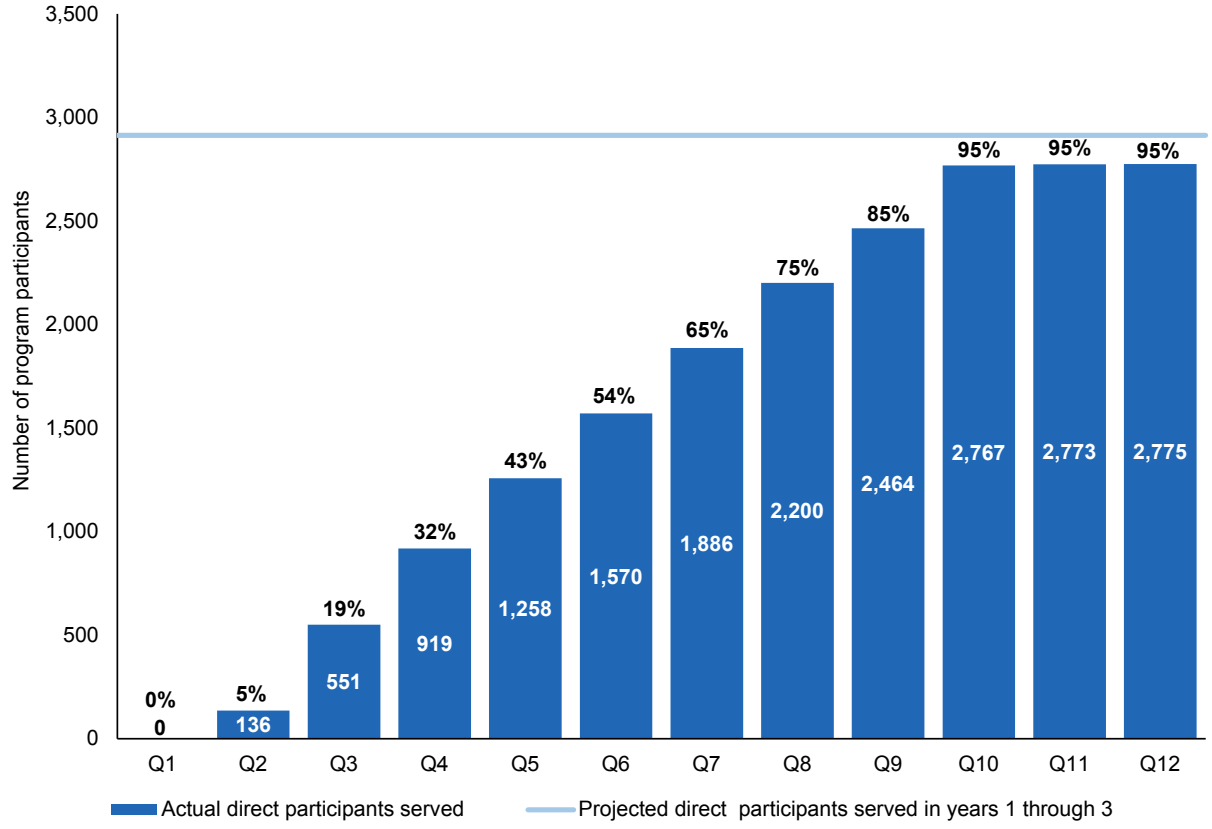
Providers referred patients to Project INSPIRE for enrollment. Care coordinators at each health system enrolled eligible patients in Project INSPIRE at Montefiore and Mount Sinai, using the health systems' electronic medical records (EMRs) to identify participants who were eligible for the program. DOHMH worked with care coordinators and providers to continue enrolling patients until February 28, 2017, as well as engaging community-based organizations to increase referrals. Enrollment ended in February to allow time for all participants to complete treatment.

DOHMH did not make any major changes to its enrollment strategies or to the delivery or services during the cooperative agreement.

b. Evidence of enrollment effectiveness

Enrollment began slowly: the awardee did not enroll any participants in the first program quarter and enrolled only 5 percent of its total target in quarter 2. However, enrollment rapidly progressed and by quarter 4 the awardee had enrolled 32 percent of its final target. At the beginning of Year 3, the awardee reduced its projected enrollment target from 3,200 participants to 2,914 participants. The awardee reduced the enrollment target and ended enrollment during the 10th program quarter to allow participants enough time to complete the program. Overall, the awardee reported that it enrolled 2,775 direct participants from January 2015 (when it launched its program) through August 2017, which represents about 95.2 percent of its final three-year projections (Figure II.1). However, when measured against the three-year projection set at the end of Year 1 (enough time for the awardee to gain a better understanding of the number of eligible participants but before the awardee began planning for the end of their cooperative agreement), the awardee met 86.7 percent of its projection. According to DOHMH's measurement and monitoring reports, 82 percent (1,844) of participants who were deemed ready for treatment by their HCV clinical provider actually initiated HCV treatment (Table II.2).

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. At the beginning of Year 3, the awardee reduced its projected enrollment from 3,200 to 2,914.

Table II.2. Treatment initiation rate for HCV infection

	Year 1 (January–August 2015)	Year 2	Year 3 (September 2016–August 2017)	Total
Number of patients determined to be treatment candidates	729	996	493	2218
Number of participants that initiated treatment	634	822	388	1844
Percentage of participants that initiated treatment	87%	83%	79%	83%

Source: Awardee self-measurement and monitoring reports.

c. Barriers and facilitators associated with enrollment effectiveness

DOHMH experienced enrollment delays in the first year due to the institutional review board (IRB) approval process and hiring delays (see our first annual report). Although the awardee overcame those barriers, it continued to struggle with the challenges of enrolling a hard-to-reach target population. Factors that facilitated the awardee's progress toward its three-year enrollment goal were ensuring provider engagement, using EMR software alerts, and reaching out to community partners working with HCV populations to increase referrals.

As reported in the first annual report, obtaining IRB approval took longer than expected at Mount Sinai because the IRB did not grant an exemption from full review as the IRBs at other partner organizations did. In addition, identifying and hiring suitable candidates for the care coordinator and peer navigator positions was challenging due to high demand for health care staff in New York City.

In Year 2, Project INSPIRE's target population became harder to reach because initial participants who were ready and waiting for treatment had already completed the program. To meet participant needs, DOHMH and its implementing sites made several adjustments to facilitate enrollment. For example, Project INSPIRE care coordinators and providers began recruiting patients at clinics that the population often visits, such as methadone clinics. However, they noted that most patients pick up their methadone in the morning and leave, not staying long enough to enroll in the program. In response to this issue, and in order to increase enrollment, the implementing sites changed their physicians' hours in the methadone clinics so that more providers were available in the morning. Care coordinators also sent "Happy New Year" cards to eligible patients who had not enrolled by the final year of the program. These holiday cards facilitated enrollment by showing the program's dedication to treating eligible patients, and the cards led to approximately 10 new enrollees.

Provider encouragement was an important facilitator of enrollment. Because of prior HCV medication side effects, patients were hesitant to accept HCV treatment. Providers played an important role in enrollment by explaining the new treatment and assuaging the patients' fears. Mount Sinai project leaders estimated that 75 to 80 percent of eligible patients whom they approached enrolled in Project INSPIRE. This may be attributed to the fact that providers presented Project INSPIRE to potential participants as a support system with no burden on patients.

Providers also worked with care coordinators to identify and enroll eligible patients. In key informant interviews, program staff discussed the value of engaging with various types of clinics. Staff felt that teaching clinics were particularly receptive to engagement and recruitment, as clinic staff were "open to knowledge" and "want to meet the patients where they're at." Key informants also reported that engagement with liver-specialty clinics facilitated recruitment, as providers had more familiarity with the program and the needs of HCV patients. In interviews, staff referenced the valuable role played by committed providers across organizations in facilitating recruitment. They discussed how providers spoke highly of the program, thereby prompting other providers to join and refer patients.

Montefiore and Mount Sinai used their EMRs to facilitate the identification of potential participants. One implementing site used proprietary software to alert providers of patients

eligible for Project INSPIRE. In addition, Montefiore attempted to identify new cases through an EMR a prompt that appeared for patients born between 1945 and 1965 who have never had an HCV test. The prompt asked the provider to order the HCV test.

Last, by using existing HCV community networks, DOHMH engaged local community stakeholders to establish partnerships and ultimately increase referrals. The implementing sites worked with the awardee leadership to place Project INSPIRE champions at primary care clinics that do not treat HCV in order to identify potential participants and increase enrollment. In addition to (1) doing a monthly chart review for patients identified at each clinical site as having HCV and (2) meeting with community-based organizations, the implementing sites' project leaders worked to identify clinical sites not treating for HCV.

DOHMH attended multiple New York City Hep C task force meetings. The task force is made up of organizations addressing HCV including, syringe exchange programs, substance abuse programs, and other community-based organizations. The task force is made up of smaller groups of organizations from all parts of the city, which meet regularly. DOHMH attended the citywide task force meetings as well as the Bronx and Harlem Hep C task force meetings. During the meetings, DOHMH provided updates on Project INSPIRE and encouraged participating community organizations to refer patients for treatment. To make it easier for the task force organizations to refer patients to Project INSPIRE, DOHMH gave them the contact information for project directors at each implementing site.

2. Delivery of program services

a. Description of and changes to service delivery model

Project INSPIRE was a new program created through HCIA R2. DOHMH intended to adapt the HIV care model to coordinating care for HCV patients.[5] DOHMH modeled Project INSPIRE after a prior care coordination program, Check Hep C, which was implemented in federally qualified health centers (FQHCs) in New York City. However, Check Hep C did not establish referral networks needed to holistically treat HCV and comorbid conditions. Project INSPIRE was implemented at two major medical centers that could treat HCV, mental health, and alcohol or drug additions. According to an analysis conducted by Weill Cornell University, Project INSPIRE was implemented with a high level of fidelity. DOHMH reported that this was a result of a well-formulated plan that incorporated an evidence-based HIV care management model.

Care coordinators and peer navigators guided participants through treatment. Peer navigators, who assist the care coordinators, are often from the community they serve and may have previously had HCV and received treatment.

The purpose of Project INSPIRE was to (1) increase number of participants in HCV treatment and the number of participants achieving an SVR, (2) improve patient satisfaction, and (3) increase alcohol or drug dependence screening and treatment for participants in order to reduce new episodes and increase treatment adherence. Project INSPIRE leaders intended to address these outcomes by using a care coordination model as follows:

[5] See https://www.hiv.gov/federal-response/policies-issues/hiv-aids-care-continuum.

- Care coordinators at both clinical partners administered a psychosocial assessment to identify participants who meet established thresholds for psychosocial intervention.
- Care coordinators then administered the appropriate health promotion modules. Participants received a multitude of care coordination services, including setting up and reminding participants about medical appointments, offering coaching and services promoting participant health, establishing linkages to clinical and non-clinical providers, and providing other forms of social and instrumental support that helped participants adhere to treatment.
- Once coordinators and physicians determined that participants were ready to initiate treatment, care coordinators completed insurance authorizations and related paperwork for as part of the prior authorization process.

After participants completed treatment, care coordinators made reminder calls to encourage them to return for SVR testing 12 weeks after treatment.

b. Evidence of service delivery effectiveness

DOHMH achieved a high level of service delivery effectiveness in implementing Project INSPIRE. We based our assessment on the delivery of intervention services, staffing and training, engagement of provider and provider organization, and engagement of participants. The awardee successfully partnered with and acquired staff at both implementing sites to deliver intervention services. DOHMH also implemented strategies to engage providers and participants throughout the cooperative agreement. We provide details below.

Delivery of intervention services. DOHMH and the implementing sites successfully delivered both care coordination and tele-mentoring intervention services. Using an evidence-based HIV care management model reduced the need for adaptations to the project plan after implementation began. The few adaptions included minimal revisions to the evidence-based health promotion material used by the care coordinators.

Based on DOHMH's measurement and monitoring report, it achieved many of the key processes identified in its theory of action, including the following:

- Ninety-three percent (1,811) of participants who initiated treatment between January 2015 and May 2017 completed treatment.
- Ninety-seven percent (1,692) of participants who completed treatment were "treatment adherent," which is defined as taking greater than 80 percent of prescribed medication.
- Eighty-six percent (730) of non-cirrhotic participants and 89 percent (820) of cirrhotic participants completed treatment and achieved an SVR. For most participants who did not achieve SVR, either their test results were still pending or they had not returned for their SVR testing appointments.

Staffing and training. Over the course of the cooperative agreement, DOHMH supported 47.3 full-time equivalent (FTE) staff to work on Project INSPIRE, including physicians, behavioral health specialists, care coordinators, peer counselors, and administrative staff. Staffing data is presented in Table II.3 below. In key informant interviews, staff opinions diverged on the manageability of the project workload. Though some interviewees reported that

the workload was manageable, others reported staffing shortages because of medical and family leave. To address staffing issues, care coordinators floated between the implementing clinical sites to ensure coverage during staff absences. According to the staff survey, approximately 37 percent of respondents considered turnover as a barrier to Project INSPIRE achieving its goals. However, in interviews at the end of Year 3, program staff did not feel as if turnover disrupted the workflow. Interviewees discussed turnover at the awardee's leadership level. In Year 2, the principal investigator left DOHMH. The project director assumed responsibility for this role in addition to the role of project director. Interviewees reported that the transition was seamless.

Table II.3. Overview of awardee staffing, by end of 12-month period

	Quarter 4	Quarter 8	Quarter 12
Total staff FTEs	37.5	40.7	47.3
Staffing as a percentage of target	79.8%	82.5%	99.8%
Number of separations	0	2	1
Retention rate	100	77.8%	87.5%

Source: Based on data reported by the awardee to the implementation and monitoring contractor.
FTEs = full-time equivalents.

Staff perceived training from DOHMH as effective. Care coordinators and peer navigators attended orientation training during the project's first year. Eighty-seven percent of respondents to the staff survey reported receiving formal training related to Project INSPIRE. Of those who indicated receiving training, nearly all (90 percent) felt it equipped them with new skills that were important for their role in the project. Further, 89 percent of respondents felt the training helped to improve their job performance. DOHMH offered informal training throughout the cooperative agreement. Nearly all respondents to the staff survey reported that they participated in staff meetings (96 percent), individual supervision (86 percent), and self-study (86 percent). A majority of respondents also reported asking a colleague for help (83 percent), mentoring (76 percent), and group supervision (71 percent).

Engagement of providers. Both the Montefiore and Mount Sinai health systems engaged providers by facilitating weekly training and the exchange of information between hepatologists and program providers through tele-mentoring. Tele-mentoring increased the familiarity of primary care providers at both clinical sites with HCV, treatment, and the population. Tele-mentoring also offered continuing medical education credits, which facilitated provider engagement and the certification of primary care providers as approved prescribers of HCV drug therapy in New York State. In Year 2, DOHMH and both clinical partners reported that primary care providers were treating more participants with HCV and leading tele-mentoring sessions. Tele-mentoring sessions evolved from teaching sessions to case conferencing, including presentations and group discussions of HCV cases.

Engagement of program participants. The main component of Project INSPIRE was care coordination, which entailed care coordinators engaging participants to enroll in the program, initiate treatment, and complete treatment. Once a patient was enrolled, care coordinators performed outreach to keep them engaged in treatment. Care coordinators called participants, sent letters, provided MetroCards for transportation, and developed relationships with them. Participants lost to follow-up often had competing needs, such as health or social needs. In the

staff survey, a majority of staff reported that Project INSPIRE successfully engaged participants in the program. Survey responses also suggested that staff resistance to the program was a major (22 percent) or minor barrier (26 percent) to achieving program goals.

Participant engagement became more challenging as care coordinator caseloads grew. In interviews, key informants reported that care coordinators had full caseloads, which may not have allowed for the desired amount of participant engagement. Thirty-two percent of respondents to the staff survey reported that staffing was a minor barrier, and 5 percent reported that staffing was a major barrier to program effectiveness. Because of their full caseloads, care coordinators were limited in other capabilities. One key informant noted that once coordinators reach the originally planned 125-participant caseload, they did not have enough time to do other work related to participant engagement.

c. Barriers and facilitators associated with service delivery effectiveness

DOHMH and its implementing sites effectively implemented and delivered services. The awardee overcame potential barriers to service delivery, including staff turnover and a lack of participant retention. Provider engagement and participant engagement were key factors for service delivery effectiveness and for overcoming these potential barriers. Other facilitators included on-site champions, provider tele-mentoring, and innovative strategies to keep participants engaged.

The possibility of curing a disease facilitated provider involvement. Project leadership reported that providers "really love" the idea of curing a deadly disease, something providers may not often have the chance to do. This aspect was so exciting for some providers that three joined the tele-mentoring sessions even though HCIA R2 funding did not cover their time. Payment for providers' time also facilitated engagement.

> "[Tele-mentoring's] become a much more peer-to-peer learning network rather than a top-down or expert-teaching-the-primary-care-docs kind of forum."
>
> —Project leader

In addition, providers value the program because it pays for part of their time. This has increased the likelihood of provider involvement by allowing providers to join tele-mentoring sessions and engage with the intervention.

Last, tele-mentoring and support facilitated broader recruitment and care provision than did face-to-face meetings Having a champion on site to provide guidance to less experienced providers opened helpful lines of communications between newer providers and champions. Project INSPIRE's tele-mentoring training taught providers how to treat HCV, skills not readily available elsewhere. The support and funding provided through Project INSPIRE enabled providers who would otherwise be too busy to spend time learning how to treat patients with HCV. The Montefiore project manager noted the potential of this model to impact the way care is provided for other diseases as well.

> "I think in that way providers are easily connected to the project and easily engaged because the average provider is just overwhelmed with the amount of work they have to do. And so, if you have a program that's helping pay just to have some of that time blocked off, they're more than happy to work on whatever is being blocked off."
>
> —Implementing site staff

Peer navigators facilitated participant engagement with an apprehensive population by sharing similar experiences, encouraging the participants to stay engaged, and meeting participants' needs. Navigators often accompanied participants to appointments, provided MetroCards, and placed reminder calls about medication adherence. However, for future implementation, key informants reported that peer navigators should have a more defined role that is integrated into the workforce. Defining a peer navigator's role may include improving their skills to better help participants achieve their goals.

> "[Participants often work multiple jobs.] That's why some of our services were delivered over the phone. And the [care coordinators and participants] would just try and find the time mutually—a time that would work for everyone so that they could at least review some of the key elements of why it's so important to take this medication—even if it was quickly."
>
> —Implementing site staff

Last, a barrier to service delivery effectiveness is that some participants may not have seen the value in care coordination. Participants working multiple jobs want to "get in and get out" of their provider's office. In response, care coordinator sand peer navigators could provide some health modules over the phone.

C. Assessment of perceived program effects on the delivery of care and outcomes

Overall, staff perceived the effects of Project INSPIRE in a positive light (see Table II.4 for staff perceptions on program impact). In the survey of program staff, 83 percent of respondents felt that the program had a positive impact on the quality of care and services provided to participants. Similar percentages of respondents felt that Project INSPIRE had a positive impact not only on the ability of staff to respond in a timely way to participant needs (83 percent) but also on the efficiency with which care or services were provided to participants (82 percent). Nearly all respondents also said that the program had a positive impact on the achievement of participant's health goals (91 percent) and care coordination (87 percent).

In key informant interviews, program staff across the implementing sites discussed the impact of the program on participants and on the attainability of project goals. They referenced the impact of care coordinators on reducing service utilization while they were discussing examples of coordinators fielding patient questions and preventing otherwise needless ED or urgent care visits. Key informants also discussed increasing patient engagement in primary care following participation in Project INSPIRE. Staff did caution, however, that the program's effects may be limited by its target populations, vulnerable individuals who are more difficult to engage and those for whom effects can be identified.

The staff who were interviewed perceived other potential positive impacts of the program. One key informant reported that the program improved participants' social well-being. Participants achieve a feeling of accomplishment following treatment completion, and the informant suggested that the "humanizing" of care provided through Project INSPIRE motivates patients and improves their self-esteem.

Table II.4. Staff perceptions of Project INSPIRE program effects on care

Percentage of non-clinician staff indicating Project INSPIRE had a positive impact on the following:	Number of respondents (N=23)	Percentage of respondents
The quality of care and services you provide to participants	19	83
Your ability to respond in a timely way to participant needs	19	83
The efficiency of care services provided to participants[a]	18	78
Your ability to provide care or services that are responsive to participant preferences, needs, and values	18	78
Care services that are provided fairly to all participants[a]	16	73
Access to care or services for all participants	19	83
Achievement of participants' health goals	21	91
Participant satisfaction	21	91
Participant quality of life	21	91
Care coordination	20	87

Source: HCIA R2 evaluation survey of participating non-clinician staff, fall 2016.

[a]One missing response.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

Project INSPIRE targeted HCV patients who often fall out of care or have never been in care. The goal is to facilitate HCV treatment by addressing comorbid conditions and enabling the initiation, adherence to, and completion of treatment. Thus, we can expect to see effects among participants who complete the program. Since HCV leads to cirrhosis of the liver and eventual hospitalization, the intervention should prevent hospitalizations by curing participants of HCV. On average, the HCV treatment regimens for participants lasted 12 weeks. Although the SVR is not documented until a minimum of 12 weeks post-treatment, DOHMH expected most participants' health to improve well before they returned for SVR testing.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of DOHMH's Project INSPIRE. We also present our preliminary findings on the baseline characteristics of the treatment group, which we identified from Medicare/Medicaid claims and enrollment data.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: DOHMH

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	350[a]
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018	2,384[a]
Number of participants needed to detect 10% effect	
Total expenditures	5,032
Likelihood of all-cause hospitalizations	1,710
Number of participants needed to detect 20% effect	
Total expenditures	1,258
Likelihood of all-cause hospitalizations	428
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	No serious issues. Proceeding with comparison group selection
Do claims identify the primary expected effects?	Some effects observed in claims data but important effects likely missing
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician and beneficiary survey
Implementation data that will be analyzed	Data on services used; sustained viral response data

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We anticipate conducting a rigorous impact analysis of Project INSPIRE. We are in the final stages of developing a valid comparison group for Medicare beneficiaries by using propensity score matching to select comparison beneficiaries who live in New York City, have HCV, and have similar sociodemographic characteristics and service use as program participants. We plan to use a similar process to select a comparison group for Medicaid beneficiaries. The projected final Medicaid and Medicare analysis samples (shown in Table III.1) are based on enrollment by February 2017, when enrollment ended. The analysis sample for Medicaid beneficiaries should be large enough to detect plausible effects on most claims-based measures; although the Medicare analysis sample is smaller, it should be large enough to detect plausible effects on some claims-based measures.

B. Characteristics of Medicare participants at baseline

This section includes a summary of both core and awardee-specific claims-based outcomes at baseline. For the purpose of our evaluation, the treatment group consists of beneficiaries who are in FFS and who were enrolled in Project INSPIRE (according to lists from the awardee) at some point between its start in January 2015 through February 28, 2017. Medicare managed care beneficiaries were excluded from the analysis because data on these beneficiaries were not available.

DOHMH began to enroll Medicare and Medicaid beneficiaries into Project INSPIRE in January 2015. As of the end of February 2017, the program had 2,775 participants.[6]

In presenting the baseline characteristics for Medicare FFS beneficiaries, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they had to have been enrolled in the awardee's program on or before February 28, 2017, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. After we excluded patients who did not meet the above criteria, 327 Medicare FFS participants (out of 1,006 total Medicare participants—FFS and managed care—enrolled through February 28, 2017) were included in the analysis of baseline characteristics for this report. The vast majority of Medicare beneficiaries who were excluded from the analysis were in Medicare managed care during their month of enrollment.

[6] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

The Medicare FFS beneficiaries that participated in Project INSPIRE are fairly typical of patients with HCV nationwide[7] in terms of demographics and health status characteristics (Table III.2). The most common characteristics of participants include being younger than 65 (41 percent); male (68 percent); black (49 percent); originally entitled to Medicare through disability (51 percent, relative to a national average of 24 percent); and dually eligible (56 percent, relative to a national average of 18 percent). A less common characteristic is being originally entitled to Medicare through end-stage renal disease (ESRD) (6 percent). None of the participants had enrolled in hospice. They had a mean hierarchical condition categories (HCC) risk score of 2.22 (relative to a national mean risk score of 1.00), a median risk score of 1.75, a 25th percentile risk score of 1.12, and a 75th percentile risk score of 2.97. Taken together, the scores indicate that the participants were substantially sicker than the average Medicare FFS beneficiary.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in DOHMH's program through February 28, 2017

	All participants (N = 327)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	135	41
65 to 74	161	49
75 to 84	27	8
85 and older	4	1
Gender		
Female	106	32
Male	221	68
Race		
White	105	32
Black	160	49
American Indian, Alaska Native, Asian/Pacific Island American, or other	9	3
Hispanic	47	14
Original reason for Medicare eligibility		
Old age and survivor's insurance	139	43
Disability insurance benefits	168	51
End-stage renal disease (ESRD)[a]	20	6
Hospice[b]	0	0
Medicare/Medicaid dual status, percentage dual[b]	182	56
HCC score[c]		Statistic

[7] As with our sample, black populations nationwide include high proportions of people with HCV. In addition, as with our sample, males are more likely than females to have HCV. See Armstrong, G. L., A. Wasley, E. P. Simard, G. M. McQuillan, W. L. Kuhnert, and M. J. Alter. "The Prevalence of Hepatitis C Virus Infection in the United States, 1999 through 2002." *Annals of Internal Medicine,* vol. 144, no. 10, 2006, pp. 705–714.

Table III.2 *(continued)*

Characteristics	All participants (N = 327)	
	Number	Percentage
Mean		2.22
25th percentile		1.12
Median		1.75
75th percentile		2.97

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of February 28, 2017.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the beneficiary signed up for the program. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Consistent with the high HCC scores, participants had high baseline expenditures (Table III.3). We examined the baseline cost of care by calculating average per beneficiary per month (PBPM)[8] Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $3,021 (relative to a national average among all Medicare FFS beneficiaries of $792 in 2014). The average PBPM Medicare payments were $1,584 for acute inpatient services (relative to a national average of $263); $573 for outpatient services (relative to a national average of $107); and $411 for physician services. The quarterly total expenditures trended upward over time, whereas the quarterly expenditure components have no clear pattern over time.

[8] All national data in this paragraph are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in DOHMH's program through February 28, 2017

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	327	293	305	327	327
Average Medicare expenditures PBPM[a]					
Total	3,021 (296)	2,499 (342)	2,976 (405)	3,015 (383)	3,482 (415)
Acute inpatient	1,584 (184)	1,093 (218)	1,672 (290)	1,518 (272)	1,977 (300)
Inpatient other[b]	59 (23)	100 (46)	17 (13)	97 (62)	24 (17)
Outpatient[c]	573 (129)	599 (175)	551 (111)	583 (141)	543 (129)
Physician services	411 (26)	369 (32)	412 (36)	403 (34)	448 (39)
Home health	97 (17)	103 (29)	89 (21)	78 (25)	117 (32)
Skilled nursing facility	269 (63)	219 (88)	212 (75)	296 (112)	340 (140)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	29 (6)	17 (5)	24 (11)	38 (16)	33 (10)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	1,007 (111)	742 (127)	1,046 (158)	1,012 (153)	1,175 (150)
Outpatient ED visits[e]	1,076 (159)	910 (149)	1,193 (221)	1,075 (259)	1,065 (161)
Primary care visits in any setting	7,404 (584)	7,099 (824)	7,011 (782)	6,820 (728)	8,333 (741)
Primary care visits in ambulatory settings	5,096 (436)	5,321 (691)	4,773 (619)	4,542 (401)	5,482 (537)
Specialist visits in any setting	17,189 (1,002)	15,038 (1,086)	16,771 (1,320)	17,056 (1,316)	19,126 (1,366)
Specialist visits in ambulatory settings	11,995 (775)	11,440 (822)	11,717 (943)	11,919 (818)	12,469 (828)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	42 (3)	13 (2)	17 (2)	17 (2)	20 (2)
Percentage with an outpatient ED visit[e]	44 (3)	16 (2)	19 (2)	16 (2)	18 (2)
Percentage with a 30-day readmission among all discharges from acute hospitals	23 (2)	12 (5)	28 (5)	27 (5)	21 (4)
Percentage of participants with a readmission among all participants	12 (2)	2 (1)	5 (1)	4 (1)	4 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of February 28, 2017.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

As would be expected for a relatively sick population, the participants generally had high rates of service use during the baseline year. The rate of acute care hospitalizations was 1,007 per 1,000 Medicare FFS beneficiaries per year—higher than the national average among all FFS beneficiaries of 274 per 1,000 beneficiaries per year in 2014. Forty-two percent of participants had at least one hospitalization in the baseline year.[9] Similarly, for participants in the baseline year, 23 percent of all discharges had a readmission within 30 days—higher than the national

[9] All national data in this paragraph except for ambulatory observation bed stays are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

average in 2014 for Medicare beneficiaries (18 percent per discharge). The rate of ED visits of 1,076 per 1,000 Medicare FFS beneficiaries per year was also higher than the national average in 2014 of 652 per 1,000 beneficiaries per year. Forty-four percent of participants had at least one ED visit in the baseline year. At baseline, the rate of primary care visits in any setting was 7,404 per 1,000 Medicare FFS beneficiaries per year; 69 percent occurred in an ambulatory setting. The rate of specialty care service use in any setting was 17,189 per 1,000 Medicare FFS beneficiaries per year; 70 percent occurred in an ambulatory setting. There was not a clear trend over time in service utilization. Overall, participants had higher expenditures, a higher rate of acute care hospitalizations, and a higher rate of readmissions relative to the national averages for all Medicare FFS beneficiaries. These findings indicate that there may be the potential for improving the care of participating beneficiaries.

Project INSPIRE is expected to have a fairly high proportion of participants with mental health or substance abuse problems, as these are problems commonly associated with people with HCV nationwide.[10,11] Table III.4 presents measures specific to Project INSPIRE, including two that relate to mental health and substance abuse. The rate of ED visits for mental health or substance abuse was 156 per 1,000 Medicare FFS beneficiaries per year. The rate of hospital admissions for mental health or substance abuse was 124 per 1,000 Medicare FFS beneficiaries per year.

Table III.4. Measures specific to DOHMH for Medicare FFS beneficiaries enrolled in the program through February 28, 2017

Types of utilization measures	12 months before enrollment	Utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	327	293	305	327	327
Health care utilization rates (annualized per 1,000)					
ED visits for mental health or substance abuse—primary diagnosis[a]	156 (60)	70 (38)	268 (134)	202 (117)	86 (40)
Hospital admissions for mental health or substance abuse—primary diagnosis, all hospitals[a,b]	124 (30)	98 (36)	188 (74)	101 (39)	110 (40)

[10] For information regarding substance abuse data, see "The Prevalence of Hepatitis C Virus Infection in the United States, 1999 through 2002." Available at http://annals.org/article.aspx?articleid=723191.

[11] For information regarding mental health data, see "Neuropsychiatric and Psychosocial Issues of Patients with Hepatitis C Infection: A Selective Literature Review." Available at http://hepatmon.com/?page=article&article_id=8340.

Table III.4 *(continued)*

Types of utilization measures	12 months before enrollment	Utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Other health status and utilization measures					
Percentage with cirrhosis of the liver	19 (2)	8 (2)	9 (2)	10 (2)	10 (2)
Percentage with a liver transplant	2 (1)	2 (1)	2 (1)	1 (1)	2 (1)
Percentage with hepatocellular carcinoma	4 (1)	2 (1)	2 (1)	2 (1)	2 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of February 28, 2017.

Note: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as the 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.
Standard errors are shown in parentheses.

[a]The search for mental health and substance abuse diagnoses is limited to the primary diagnosis on the claim.

[b]Unlike the acute hospital admissions measure in Table III.3, the measure for hospital admissions for mental health or substance abuse includes rehabilitation, long-term care, and psychiatric hospitals.

ED = emergency department; FFS = fee-for-service.

HCV can lead to liver problems such as cirrhosis, the need for a transplant, or hepatocellular carcinoma. Table III.4 also shows that the percentage of participants in the baseline year with a diagnosis on a Medicare claim for each of these conditions was 19 percent, 2 percent, and 4 percent, respectively.

C. Characteristics of Medicaid participants at baseline

In presenting the baseline characteristics for Medicaid beneficiaries, we included both Medicaid FFS and Medicaid managed care beneficiaries. Similar to the restrictions imposed on the Medicare FFS beneficiaries, we restricted the treatment group to Medicaid beneficiaries who were enrolled in Medicaid when their eligibility for awardee-provided services began (that is, their enrollment date) and for 90 consecutive days during the baseline year (the 365 days immediately before their enrollment). In addition, they had to have been enrolled in the awardee's program on or before June 30, 2015 to be included in this report due to lags in Medicaid data availability. After we excluded participants who did not meet the above criteria, 552 Medicaid participants (out of 1,669 total Medicaid participants enrolled through May 31, 2016) were included in the analysis of baseline characteristics for this report. About two-thirds of the program's 1,669 Medicaid participants (1,078 beneficiaries) were excluded from this analysis because they enrolled after June 30, 2015. However, these participants should be included when more recent data become available.

Baseline demographic characteristics are presented in Table III.5 and Table III.6 for beneficiaries who enrolled in the program in the first or second quarter of 2015. As shown in Table III.5, the most common characteristics of Medicaid participants include being 55 to 64 years old (48 percent), male (61 percent), Hispanic (45 percent), eligible for full Medicaid benefits (97 percent), in the eligibility category of supplemental security income (SSI) Blind/Disabled (52 percent), enrolled in a comprehensive managed care plan (68 percent), non-dual eligible (75 percent), not being enrolled through a Home and Community Based Services (HCBS) waiver (99 percent), and having no third-party insurance (99 percent).

In Table III.6, we find that the most common Chronic Disability Payment System (CDPS) categories of Medicaid participants include infectious disease (87 percent), cardiovascular (60 percent), gastrointestinal (55 percent), substance abuse (43 percent), and psychiatric disorders (43 percent). Ninety-six percent of participants are in at least one CDPS category. Although only 87 percent of participants had an infectious disease diagnosis in their claims data in the 365-day baseline period, it is anticipated that all participants will have a diagnosis of HCV in their claims data at some point in time (that is, either in the baseline period or before the baseline period), because HCV is a requirement to be in the demonstration. The high rates of substance abuse (43 percent) and psychiatric disorders (43 percent) are consistent with the high rates for these conditions found among people with HCV nationwide.[12,13] The participants have a mean risk score of 3.01, a median risk score of 2.64, a 25th percentile risk score of 1.78, and a 75th percentile risk score of 3.63.

[12] For information regarding substance abuse data, see "The Prevalence of Hepatitis C Virus Infection in the United States, 1999 through 2002." Available at http://annals.org/article.aspx?articleid=723191.

[13] For information regarding mental health data, see "Neuropsychiatric and Psychosocial Issues of Patients with Hepatitis C Infection: A Selective Literature Review." Available at http://hepatmon.com/?page=article&article_id=8340.

Table III.5. Baseline year demographic characteristics of Medicaid beneficiaries enrolled in DOHMH's program through the second program quarter (June 30, 2015)

	All enrollees (N = 552)[a]	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
18–21	1	0.18
22–34	17	3.08
35–44	34	6.16
45–54	164	29.71
55–64	266	48.19
65–74	66	11.96
75–84	4	0.73
Gender		
Female	213	38.59
Male	339	61.41
Race and ethnicity		
White	51	9.24
Black	188	34.06
Asian or Pacific Islander	7	1.27
American Indian, Alaska Native, or other	6	1.09
Hispanic	2	0.36
Hispanic and one or more races	245	44.38
More than one race (not Hispanic)	5	0.91
Type of benefits		
Full Medicaid benefits	536	97.10
Restricted benefits	16	2.90
Medicaid eligibility category		
SSI aged	34	6.16
Non-SSI aged	14	2.54
SSI blind/disabled	289	52.36
Non-SSI blind/disabled	36	6.52
TANF, safety net, or low income family adults	118	21.38
All other adults	61	11.05
Managed care enrollment		
Comprehensive managed care plan	373	67.57
Long-term care carve-out	38	6.88
No managed care enrollment	141	25.54

Table III.5 *(continued)*

Characteristics	All enrollees (N = 552)[a]	
	Number	Percentage
Medicare/Medicaid dual status, percentage dual		
Dual	140	25.36
Non-dual	412	74.64
HCBS waiver enrollment		
Enrolled in any HCBS waiver	7	1.27
Not enrolled in an HCBS waiver	545	98.73
Third-party insurance		
Third-party insurance	6	1.09
No third-party insurance	546	98.91
Quarter of initial program enrollment		
Q1 2015	208	37.68
Q2 2015	344	62.32
Records included in the expenditure and utilization analysis[b]		
	530	96.01

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of June 30, 2015.

Note: The baseline period is defined as one year (365 days) prior to each individual beneficiary's program enrollment date. Medicaid enrollees must have at least 90 days eligibility in the baseline period and on the date of program enrollment to be included in the eligible sample. All beneficiary characteristics (other than CDPS category and risk score) are measured in the last month of the baseline period.

[a]Excludes Medicaid beneficiaries who have state plan enrollment during the month of program enrollment.

[b]Excludes Medicaid beneficiaries who have state plan enrollment, partial benefits, or third-party benefits during the month of program enrollment.

HCBS = Home and Community Based Services; SSI = Supplemental Security Income; TANF = Temporary Assistance for Needy Families.

Table III.6. CDPS categories of Medicaid beneficiaries enrolled in DOHMH's program through the second program quarter (June 30, 2015)

Characteristics	All enrollees (N = 552)[a]	
	Number	Percentage
Selected CDPS category[b]		
Beneficiaries in one or more CDPS categories	531	96.2
Infectious disease	478	86.59
Cardiovascular	332	60.14
Gastrointestinal	306	55.43
Substance abuse	237	42.93
Psychiatric	235	42.57
Beneficiaries not in a CDPS category	21	3.8
CDPS risk score[b]		
Mean	3.01	
25th percentile	1.78	
Median	2.64	
75th percentile	3.63	

Source: Mathematica analysis of information from awardee's finder file and Medicaid claims and enrollment data as of June 30, 2015.

[a]Excludes Medicaid beneficiaries who have state plan enrollment during the month of program enrollment.
[b]Categories and risk scores are defined using the CDPS software and are based on the ICD-9 codes on Medicaid claims occurring during each beneficiary's one-year baseline period.
CDPS = Chronic Disability Payment System.

Baseline expenditure and utilization statistics for Medicaid participants are presented in Table III.7 and Table III.8 for non-dual status participants (enrolled in Medicaid only) and dual status participants (enrolled in Medicare and Medicaid), respectively. Unlike the beneficiary characteristics presented in Tables III.5 and III.6, the expenditure and utilization analysis presented in Tables III.7 to III.9 excludes Medicaid beneficiaries who have partial benefits or third-party benefits during the month of program enrollment (as well as those who have state plan enrollment).

In Table III.7, we find that the total average PBPM Medicaid payment among Medicaid non-dual status participants during the baseline year was $4,096. The average PBPM Medicaid payments for non-dual status participants were $687 for acute inpatient stays; $39 for ED visits; $2,390 for pharmacy services; and $979 for other services. The large proportion of total spending accounted for by pharmacy expenditures is consistent with the high cost of the drugs commonly used to treat HCV. There was not a clear pattern in quarterly expenditures over time, though expenditures did rise sharply in quarter 4, driven primarily by a spike in pharmacy spending.[14]

[14] Note that Medicare and Medicaid cover different services so Medicaid and Medicare spending patterns may be quite different. Also, while we report Medicaid prescription drug spending, due to a lack of data availability, we do not report Medicare Part D prescription drug spending.

Table III.7 also provides utilization data for the Medicaid non-dual status participants. The rate of acute hospital admissions was 676 per 1,000 Medicaid beneficiaries per year. The rate of ED visits was 1,677 per 1,000 Medicaid beneficiaries per year, with 22 percent (365 of 1,677) leading to an inpatient stay. Both acute hospital admissions and ED visits peaked in quarter 2.

Table III.7. Baseline year expenditures and health care utilization for Medicaid non-dual status beneficiaries enrolled in DOHMH's program through the second program quarter (June 30, 2015)

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees[a]	406	381	395	406	406
Average Medicaid expenditures PBPM[b]					
Total payment	4,096 (251)	3,310 (258)	3,986 (426)	3,446 (294)	5,561 (434)
Acute inpatient stays	687 (104)	573 (107)	961 (232)	585 (126)	633 (144)
Total ED payment	39 (4)	37 (6)	49 (13)	38 (5)	34 (5)
ED visits that lead to an inpatient stay	7 (2)	4 (1)	14 (9)	4 (1)	5 (1)
ED visits that do not lead to an inpatient stay	33 (4)	33 (5)	35 (10)	33 (5)	30 (4)
Pharmacy	2,390 (178)	1,795 (209)	1,880 (236)	1,880 (221)	3,931 (389)
Other[c]	979 (73)	906 (69)	1,095 (142)	944 (108)	962 (83)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions	676 (88)	661 (110)	807 (182)	690 (121)	553 (102)
Total ED visits	1,677 (157)	1,694 (210)	1,810 (509)	1,619 (187)	1,589 (199)
ED visits that lead to an inpatient stay	365 (53)	309 (62)	455 (98)	330 (78)	365 (76)
ED visits that do not lead to an inpatient stay	1,312 (134)	1,385 (189)	1,355 (494)	1,290 (162)	1,224 (158)

Source: Mathematica analysis of information from awardee's finder file and Medicaid claims and enrollment data as of June 30, 2015. New York State provided Medicaid data for services dated through June 30, 2015.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter

Table III.7 *(continued)*

according to the number of days each beneficiary is enrolled in Medicaid during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Excludes Medicaid beneficiaries who have state plan enrollment, partial benefits, or third-party benefits during the month of program enrollment.

[b]Total Medicaid expenditures for the baseline year or a given quarter exclude capitated payments. Expenditures were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[c]Other services include non-office visit practitioner services, dental care, eye care, home health care, laboratory services, intermediate care facility services, nursing home services, child care, and clinic services.

ED = emergency department; PBPM = per beneficiary per month.

In Table III.8, we find that the total average PBPM Medicaid payment among Medicaid dual status participants during the baseline year was $1,676. The average PBPM Medicaid payments for dually eligible participants were $119 for acute inpatient stays; $5 for ED visits; $305 for pharmacy services; and $1,246 for other services. Common other services received by participants include home health, laboratory, intermediate care facility, and clinic services. Spending on these service types will be broken out separately in future reports. For dually eligible beneficiaries, Medicare is typically the primary payer. Thus, the majority of spending for dual eligibles will be paid for by Medicare. In contrast, for non-dual status Medicaid beneficiaries, Medicaid is typically the primary payer. Thus, Medicaid pays for the majority of the services for non-dual status beneficiaries. As a result, the total dual status Medicaid spending ($1,676) is quite low relative to the total non-dual status Medicaid spending ($4,096). There was not a clear pattern in quarterly expenditures over time, though expenditures were highest in quarter 1 (the earliest quarter) for many service categories (total expenditures, acute inpatient stays, and pharmacy services).

Table III.8 also provides utilization data for the Medicaid dually eligible participants. The rate of acute hospital admissions was 842 per 1,000 Medicaid beneficiaries per year. The rate of ED visits was 1,577 per 1,000 Medicaid beneficiaries per year, with 19 percent (297 of 1,577) leading to an inpatient stay. For both acute hospital admissions and ED visits, the trend in service use over time increases from quarter 1 to quarter 2, decreases from quarter 2 to quarter 3, and increases from quarter 3 to quarter 4.

Table III.8. Baseline year expenditures and health care utilization for Medicaid dually eligible beneficiaries enrolled in DOHMH's program through the second program quarter (June 30, 2015)

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees[a]	124	120	120	124	124
Average Medicaid expenditures PBPM[b]					
Total payment	1,676 (295)	2,151 (585)	1,274 (187)	1,588 (438)	1,693 (403)
Acute inpatient stays	119 (25)	213 (85)	79 (15)	97 (32)	90 (22)
Total ED payment	5 (2)	6 (3)	6 (2)	4 (1)	6 (2)
ED visits that lead to an inpatient stay	0 (< 0.5)	1 (< 0.5)	1 (< 0.5)	0 (< 0.5)	0 (< 0.5)
ED visits that don't lead to an inpatient stay	5 (2)	5 (2)	5 (2)	3 (1)	6 (2)
Pharmacy	305 (106)	560 (280)	272 (113)	99 (53)	296 (192)
Other[c]	1,246 (231)	1,372 (457)	917 (147)	1,388 (433)	1,300 (287)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions	842 (122)	774 (159)	869 (158)	853 (239)	871 (186)
Total ED visits	1,577 (324)	1,346 (369)	1,971 (610)	1,247 (273)	1,742 (405)
ED visits that lead to an inpatient stay	297 (61)	303 (118)	401 (120)	230 (95)	258 (89)
ED visits that don't lead to an inpatient stay	1,280 (303)	1,043 (328)	1,570 (590)	1,017 (228)	1,484 (387)

Source: Mathematica analysis of information from awardee's finder file and Medicaid claims and enrollment data as of June 30, 2015. New York State provided Medicaid data for services dated through June 30, 2015.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter is 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and in each baseline quarter according to the number of days that each beneficiary was enrolled in Medicaid during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Excludes Medicaid beneficiaries who have state plan enrollment, partial benefits, or third-party benefits during the month of program enrollment.

[b]Total Medicaid expenditures for the baseline year or a given quarter exclude capitated payments. Expenditures were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[c]Other services include non-office visit practitioner services, dental care, eye care, home health care, laboratory services, intermediate care facility services, nursing home services, child care, and clinic services.

ED = emergency department; PBPM = per beneficiary per month.

Project INSPIRE is expected to have a fairly high proportion of participants with mental health or substance abuse problems because these are problems commonly associated with people with HCV nationwide.[15,16] Table III.9 presents measures specific to Project INSPIRE, including two that relate to mental health and substance abuse. The rate of ED visits for mental health or substance abuse was 175 per 1,000 Medicaid beneficiaries per year. The rate of hospital admissions for mental health or substance abuse was 125 per 1,000 Medicaid beneficiaries per year.

DOHMH expects the program to reduce the number of hospitalizations among those with comorbid conditions[17] by 30 percent, compared with what it would have been absent the program. As shown in Table III.9, the rate of acute care hospitalizations among those with a comorbid condition was 773 per 1,000 Medicaid beneficiaries per year. Given the relatively high rate of acute care hospitalizations among those with comorbid conditions, we believe there is potential to improve the care for this subpopulation.

Hepatitis C can lead to liver problems such as cirrhosis, liver transplantation, or hepatocellular carcinoma. Table III.9 shows that the percentage of Medicaid participants in the baseline year with a diagnosis on a Medicaid claim for each of these conditions was 25 percent, 2 percent, and 2 percent, respectively.

[15] For information regarding substance abuse data, see "The Prevalence of Hepatitis C Virus Infection in the United States, 1999 through 2002." Available at http://annals.org/article.aspx?articleid=723191.

[16] For information regarding mental health data, see "Neuropsychiatric and Psychosocial Issues of Patients with Hepatitis C Infection: A Selective Literature Review." Available at http://hepatmon.com/?page=article&article_id=8340.

[17] Comorbid conditions refer to HIV, substance abuse disorders, mental health disorders, diabetes, and heart disease.

Table III.9. Measures specific to the awardee for all Medicaid beneficiaries enrolled in DOHMH's program through the second program quarter (June 30, 2015)

Types of utilization measures	12 months before enrollment	Utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees[a]	530	501	515	530	530
Health care utilization rates (annualized per 1,000)					
ED visits for mental health or substance abuse—primary diagnosis[b]	175 (60)	121 (40)	229 (131)	161 (61)	189 (66)
Hospital admissions for mental health or substance abuse—primary diagnosis[b]	125 (44)	105 (43)	158 (96)	146 (51)	91 (32)
Hospital admissions among those with a comorbid condition[c]	773 (81)	758 (102)	885 (161)	780 (119)	669 (98)
Other health status and utilization measures					
Percentage with cirrhosis of the liver	25 (2)	11 (1)	11 (1)	11 (1)	15 (2)
Percentage with a liver transplant	2 (1)	1 (0)	1 (0)	1 (1)	1 (1)
Percentage with hepatocellular carcinoma	2 (1)	1 (0)	1 (0)	1 (1)	1 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicaid claims and enrollment data as of June 30, 2015.

Note: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as the 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in Medicaid during the baseline year and in each baseline quarter.
Standard errors are shown in parentheses.

[a]Includes both dual and non-dual Medicaid beneficiaries. Excludes Medicaid beneficiaries who have state plan enrollment, partial benefits, or third-party benefits during the month of program enrollment.

[b]The search for mental health and substance abuse diagnoses is limited to the primary diagnosis on the claim.

[c]Comorbid conditions refer to HIV, substance abuse disorders, mental health disorders, diabetes, and heart disease.

ED = emergency department.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

DOHMH and the two partnering MCOs, HealthFirst and VNSNY, drafted a payment model consisting of a one-time bundled payment from Medicare or from Medicaid MCOs, which would fund integrated care delivery services for patients with HCV. The two partnering MCOs are negotiating with Montefiore and Mount Sinai to implement the payment model. The awardee is also conducting a return-on-investment (ROI) analysis, which may help these negotiations.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

The payment model consists of a one-time bundled payment from Medicare or from Medicaid MCOs to a facility to pay for integrated care services for each patient with HCV. The bundled payment is intended to support both the physician's time spent in conferences with care coordinators and tele-mentoring, as well as the care coordinators' time. In addition to the bundled payment, the payment model has a shared savings/shared loss component. Each facility is assigned a threshold level or "target" SVR rate (for example, 80 percent of patients obtain SVR). Facilities exceeding the threshold receive a bonus payment (for each patient with SVR), and facilities falling short of the threshold have to pay back a loss (for each patient without SVR). Beneficiaries are attributed to a facility when a provider at the facility bills using select diagnosis or CPT codes related to HCV.

Treatment of HCV consists of three phases. Phase 1 includes a pre-treatment assessment to address barriers to treatment (for example, mental health and substance abuse treatment). Phase 2 is the treatment period and Phase 3 is the report of SVR. Care coordinators are involved in all phases of the intervention. The awardee calculated that, based on its enrollment through February 2017, the cost of providing these services for all three phases combined would be $760 per beneficiary on average.

Although care coordination services are offered within the Medicaid Health Home model, the awardee stated that HCV is not one of the eligible chronic conditions for that model. As a result, some HCV patients are not eligible. (It should be noted that many HCV patients are eligible because they have chronic conditions in addition to HCV.) Furthermore, DOHMH considers the integrated care services offered by Project INSPIRE to be much more intensive than those provided in the Health Home model.

C. Status of the payment model

HealthFirst and VNSNY are currently negotiating with the two participating health systems, Montefiore and Mount Sinai. DOHMH is not involved in the negotiations because it is not a medical provider. However, the awardee is conducting a ROI analysis, which may be useful in

the negotiations between the payers and the health systems. The awardee expects to have the analysis complete by December 2018.

Currently, community health workers are funded by HCIA R2. The awardee is also trying to get a state plan amendment put in place that would help pay for and credential these workers.

Last, the awardee is exploring the use of the federal 340B program to help fund HCV care coordination services. The program helps organizations obtain outpatient drugs at a reduced price. The awardee would like for the hospital's savings from purchasing drugs at a discounted rate through the 340B program to be used to fund HCV integrated care delivery services.

D. Factors associated with the development of the payment model

In developing the payment model, the awardee has worked with numerous partners who provided input and advice. HealthFirst and VNSNY helped in researching the various payment model alternatives. Both organizations appreciated being involved in the development of the payment model, as opposed to being presented with it after it was complete. The awardee reported that one managed care organization is eager to finish negotiations and sign a contract, whereas the other organization is waiting for the awardee's claims analysis of outcomes. According to DOHMH, one of the major barriers in developing the payment model was the time needed to accumulate outcomes data.

Other partners who provided input include the following:

- Physicians provided valuable input regarding the type of treatment, length of treatment, and phases of treatment.
- Another HCIA R2 awardee is working with DOHMH on developing a state plan amendment that would help pay for and credential community health workers.
- Officials involved with the Medicaid redesign program from the New York State Department of Health also provided feedback on the payment model. However, the awardee has said that the state already has an existing Medicaid redesign process in place, making it difficult (or too late) for payment policy related to HCV treatment to be included in the process.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

DOHMH had made considerable progress on planning for sustainability. While awaiting full acceptance and implementation of its payment model, the awardee and its hospital partners gained a commitment and resources from their leadership to continue the program in the short-term and were trying to obtain additional funding. At the same time, they modified the program slightly to contain its costs in some ways while making it more robust in other ways. The program has not been scaled or replicated elsewhere, although the awardee and its partners believe that it could be, and they have pursued some activities toward that end.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, DOHMH began to aggressively look for strategies to sustain its Project INSPIRE program. The awardee worked to garner buy-in for the program from leaders at Montefiore and Mount Sinai, and at HealthFirst and VNSNY by presenting data on the program's impacts on costs and clinical outcomes. As mentioned, the awardee developed and tested a payment model to pitch to these leaders and explored integrating incentive payments into the model to fund the program's tele-mentoring component. DOHMH plans to eventually present the program to other

local medical systems, but it had no near-term, concrete plans to scale or replicate the program by the end of Year 3.

C. Implementing the SSR plan: progress and changes

Sustainability. In Year 3, DOHMH made considerable progress in developing a payment model, a major strategy for sustaining its program. As described in Chapter IV, the awardee had a multi-pronged approach to receiving ongoing payments, with the degree of progress varying by payer in descending order from Medicare, Medicaid, a potential State Plan Amendment, and the federal 340B prescription drug program.

However, because of the lag in receiving data on the program's impact, which is needed to finalize the contracts with payers, the awardee expects to need other funding for at least a year after the cooperative agreement ends. DOHMH received a six-month no-cost extension, which it planned to use not only to complete activities related to the program evaluation but also to continue to build support for the payment model, secure additional funding, and publish materials about the program to foster SSR.

DOHMH and its partners pursued other funding sources to carry the program in the short term. The awardee applied for grants through Pharma-related foundations and the city council. Although the city had traditionally provided funding for its public health initiatives, the awardee was not successful in getting local funding for the program. The two hospital partners established ways to sustain the program at their sites: Montefiore secured funding from its executive leadership team, and Mt. Sinai planned to sustain the program until 2019 with funding from its Liver Institute.

To help with longer-term sustainability, the hospitals were determining whether they could link the program to the state's medical home program just for HCV patients in order to receive the PBPM payment. The awardee reported that it already performs many functions of the medical home program (for example, a case manager follows the patient, provides appointment reminders, and connects the patient to other services). Montefiore was determining how many patients might be eligible (those with two or more chronic conditions or a mental health condition).

At the same time, DOHMH modified the program by scaling it back in other ways to help control the costs while making it more robust. Many program staff left as the grant period came to an end. Mt. Sinai lost a project manager, a social worker, a data manager, and four care coordinators. All remaining positions except for the project manager were filled. Mt. Sinai was working to fill the project manager position when the cooperative agreement ended. Montefiore secured enough funding to sustain a project manager and three care coordinators. With a smaller program staff, the awardee was integrating program components and activities into existing staff roles: the health educator employees at the clinical sites would pick up the health promotion piece, and the psychosocial screening carried out by the care coordinators will transition to existing clinical staff. By spreading these functions, the awardee expected a smaller group of remaining program staff to work more intensely on care coordination activities and pre-authorizations unique to HCV patients.

Montefiore expected to make some changes to the program but recognized the need to strategize and think about how individual changes could affect the whole program. It was considering changing the care coordination protocol and the clinical sites at which it places the care coordinators. The hospital was also reviewing the data to understand how demand for the program was changing over time across sites of care so that it could adjust the frequency (up or down) with which the care coordinators visit the sites. The hospital also planned to start screening every enrollee for depression, substance abuse, and anxiety. In addition, program staff found that the remaining patients might be the most difficult to engage, so program staff would need to work more with community-based organizations on how to improve services and peer education. In addition, Montefiore planned to downsize tele-mentoring both because the current manager was deployed elsewhere as HCIA R2 funding ended and because there were fewer primary care physicians left to train. Montefiore expected tele-mentoring to become a low-cost, low-effort virtual consulting platform that providers use if they have questions.

Mount Sinai expected to scale back its program in a few ways. Staff would deliver health promotion modules only as needed (that is, if the issue was present in the patient's risk history). The hospital would retain the Access© database, but it would use just one data entry tool. Mount Sinai would also provide a direct connection between the program and the hospital system's linkage to the care program via a direct handoff from outpatient navigators to the care coordination team.

Scalability. It did not appear that DOHMH itself planned to scale the program, but some of the individual sites reported such plans. Montefiore was thinking about whether and how to scale the program and expected that these internal discussions would intensify. Mount Sinai was considering expanding the program to privately insured patients if it had sufficient staff capacity.

Replicability. DOHMH expected the program to be readily replicable at non-academic medical center hospitals and community health centers, and potentially at large physician group practices, but it thought that more time was needed to identify specific places that are ripe for replication. The awardee's payer partners expressed interest in replicating the program in other clinical sites with which they contract. The awardee planned to write up the findings, submit them to peer-reviewed journals, and present them at conferences, but this was slow going because the awardee has few researchers on staff.

D. Factors associated with progress toward implementing the SSR plan

In addition to the factors related to demonstrating the program's value and securing a payment model (Chapter IV), in key informant interviews and the staff survey respondents shared a few additional barriers and facilitators to SSR. Mount Sinai was able to build strong internal support for the program in three main ways: (1) making sure everyone knew who the program team (especially the care coordinators) was and what it was there to provide, (2) having a hospital leader champion who communicated the value of care coordination, and (3) having nurses and providers who vocalized their concerns about their ability to care adequately for patients if the program ended. If they could not obtain sufficient funding after the cooperative agreement, Montefiore respondents had concerns about their ability to maintain the program without disruption because of the significant time needed to rehire the care coordinators who left (because of uncertainty about their jobs after the cooperative agreement ended) and to work out

the contracts, hiring, and other logistics once the hospital or another source provided new funding.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension that allows it to continue to enroll new beneficiaries. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in DOHMH's Project INSPIRE by using all enrollees with available data. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicare and Medicaid beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey to CMS the results of the match, separately for Medicare and Medicaid participants, including a balance chart, as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Four Seasons Compassion for Life

April 18, 2018

Craig Schneider (Mathematica Policy Research)
Rumin Sarwar (Mathematica Policy Research)
Laura Kimmey (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich

Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Four Seasons Compassion for Life, a nonprofit hospice and palliative care organization based in western North Carolina, received HCIA R2 funding to expand the Increasing Patient

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Four Seasons received a no-cost extension through November 30, 2017. They enrolled participants through August 31, 2017.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

and System Value with Community-Based Palliative Care (CPC) program to other providers and nearby communities. Palliative care received at least one year before the death of a patient with a life-limiting illness can improve the patient's quality of life and reduce the cost of health care. However, these patients may not seek palliative care services perhaps because they and their providers may not be aware of the benefits of palliative care and its relationship to terminal illness, or they may have misconceptions about it. Patients with advanced illnesses may not seek palliative care for other reasons as well. They often see multiple providers who do not necessarily coordinate care; they receive expensive treatments that are of limited effectiveness and that may cause detrimental side effects; and they need psychosocial or spiritual support in addition to medical care. These factors can contribute to higher utilization, increased costs, and additional suffering.

The CPC program provided patient-centered palliative care to participants with life-limiting illnesses through a highly collaborative, multidisciplinary care team that served participants' needs holistically; that is, the team provided spiritual and social support as well as clinical care. Services focused on achieving the participants' goals related to symptom management, quality of life, psychosocial and spiritual support, coordination with community-based resources, and advance care planning. HCIA R2 funding also supported the program's activities to educate participants, families, and providers about palliative care.

The HCIA R2–funded CPC program was launched on September 2, 2014, at two sites: Four Seasons and Palliative Care and Hospice of Catawba Valley (PCHCV). Four Seasons had implemented a similar version of the program prior to the cooperative agreement. During the first two years of implementation under HCIA R2 funding, Four Seasons expanded its service area to three sites: one in Asheville, North Carolina, and two in Greenville, South Carolina. Four Seasons also collaborated with the Duke Clinical Research Institute to collect data, conduct an internal evaluation, and develop a payment model for the CPC program. See Table I.1 for more details on the program's characteristics.

Four Seasons' goals were to (1) reduce hospitalizations by 10 percent, (2) reduce in-hospital deaths by 15 percent, and (3) save over $25 million during the three-year cooperative agreement.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Four Seasons Compassion for Life (FSCL) enrolls patients with life-limiting illnesses in the CPC program and provides them with a continuum of services that focus on integrating care and addressing participant needs. FSCL also is seeking to change the behavior of participants and physicians by educating participants and their families, providers, and communities about palliative care.
Major innovation	To implement in other health care organizations and regions a model of community-based palliative care in inpatient and outpatient settings.
Program components	• Integrated care • Education and training
Target population	Individuals older than 65 who are enrolled in traditional Medicare and who have a life-limiting illness, usually with a prognosis of three years or less. (The HCIA R2 funding supported education for program staff, program clinicians, external clinicians, and the community, but it did not directly fund care delivered to participants.)

Table I.1 *(continued)*

Program characteristic	Description
Theory of change/ theory of action	If a continuum of services addresses participants' needs and integrates care in all the settings through which participants with advanced illnesses transition, then the participants' outcomes will improve, and Medicare costs will be reduced. If participants, families, providers, and communities are educated in palliative care, then the behavior of participants and physicians will change such that the use of community-based palliative care will increase.
Payment model	New fee-for-service (FFS) payment, bundled or episode payment
Award amount	$9,569,123
Effective launch date	September 2, 2014
Program setting	Any setting in which a participant receives health care, including specialty care clinics, hospitals, long-term care facilities, hospices, primary care practices, or a participant's private residence
Market area	Rural, suburban, urban
Market location	Western NC and Greenville, SC
Target outcomes	• 10 percent reduction in hospitalizations for CPC participants • 15 percent fewer hospital deaths among CPC participants • $25,272,000 in total savings on the cost of care for participants who receive the CPC intervention during the three-year cooperative agreement

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that Four Seasons was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on six factors. First, the awardee enrolled 5,803 participants—108 percent of its enrollment target—by the end of the initial cooperative agreement.[4] Second, results from surveys and interviews indicated that program staff were able to deliver program services and engage participants as intended, although the lack of awardee-reported service delivery measures limited our ability to assess service delivery effectiveness. Third, the awardee was able to hire and retain program staff throughout the cooperative agreement, despite challenges with workforce shortages related to the geographic area served and burnout among clinicians who delivered palliative care services. Fourth, the awardee successfully trained program staff and clinicians to operate the program and deliver program services. Fifth, the awardee successfully engaged providers and provider organizations to refer participants and implement the program, respectively. Finally, almost all participating clinicians and other implementation staff reported that the program had a positive effect on the delivery of care.

Impact evaluation. Due to the lack of a strong comparison group, we do not anticipate being able to conduct a rigorous impact analysis for Four Seasons. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

[4] Four Seasons set a revised three-year enrollment target of 5,371 in the 9th program quarter, a 33 percent reduction from its original enrollment target of 8,000, set in the first program quarter.

Payment model. The late arrival of claims data delayed the development and implementation of the awardee's payment model. Four Seasons will, however, continue to develop and refine the model by working with other palliative care organizations and national leaders. Building on the data and the experience it gained from the program, the awardee has partnered with the American Academy of Hospice and Palliative Medicine (AAHPM) to develop a hospice- and home care–based alternative bundled payment model.

Sustainability plans. Four Seasons and most of its implementing sites reported that they have plans to sustain the CPC program. Some sites will also expand the program after the cooperative agreement to new populations or by providing new services. The implementing sites said that they found the program to be valuable and that it aligned with their organizations' priorities, making the decision to sustain and expand the program relatively easy. However, none of the sites has enough revenue for palliative care services, posing a challenge to sustainability for all of them and possibly preventing one site from sustaining the program at all. Four Seasons also began finalizing agreements with an organization that is interested in replicating the program in North Carolina.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded around the start of the third program year with a sample of 25 potential respondents and achieved a response rate of 100 percent. The clinician survey was fielded in the second half of the third program year with a sample of 25 potential respondents and achieved a response rate of 92 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

Four Seasons developed a paper screening tool and gave it to referring physicians to help them identify potential participants. CPC program staff enrolled patients at any point throughout the continuum of care if they were at least 65 years old and enrolled in traditional fee-for-service (FFS) Medicare. They also had to have a life-limiting illness[5] with a prognosis of one year or less. Referrals to the program could be initiated by a primary care physician, a hospital or other medical facility, or by patients themselves. After enrolling eligible patients, program staff

[5] One commonly accepted definition of life-limiting illness is the Center to Advance Palliative Care's (CAPC) definition: "any disease/disorder/condition that is known to be life-limiting (for example, dementia, cardiac disease, cancer) or that has a high chance of leading to death." CAPC is a national, multi-stakeholder organization created by the Robert Wood Johnson Foundation.

notified them by letter that they were enrolled in the CPC program. The letter did not ask for the participants' consent. Participants remained enrolled in the program until any of the following happened: they were discharged, they transitioned to a hospice, or they died. Participants could be discharged if they no longer needed the program medically, if they met their care plan goals, or if they asked to be discharged.

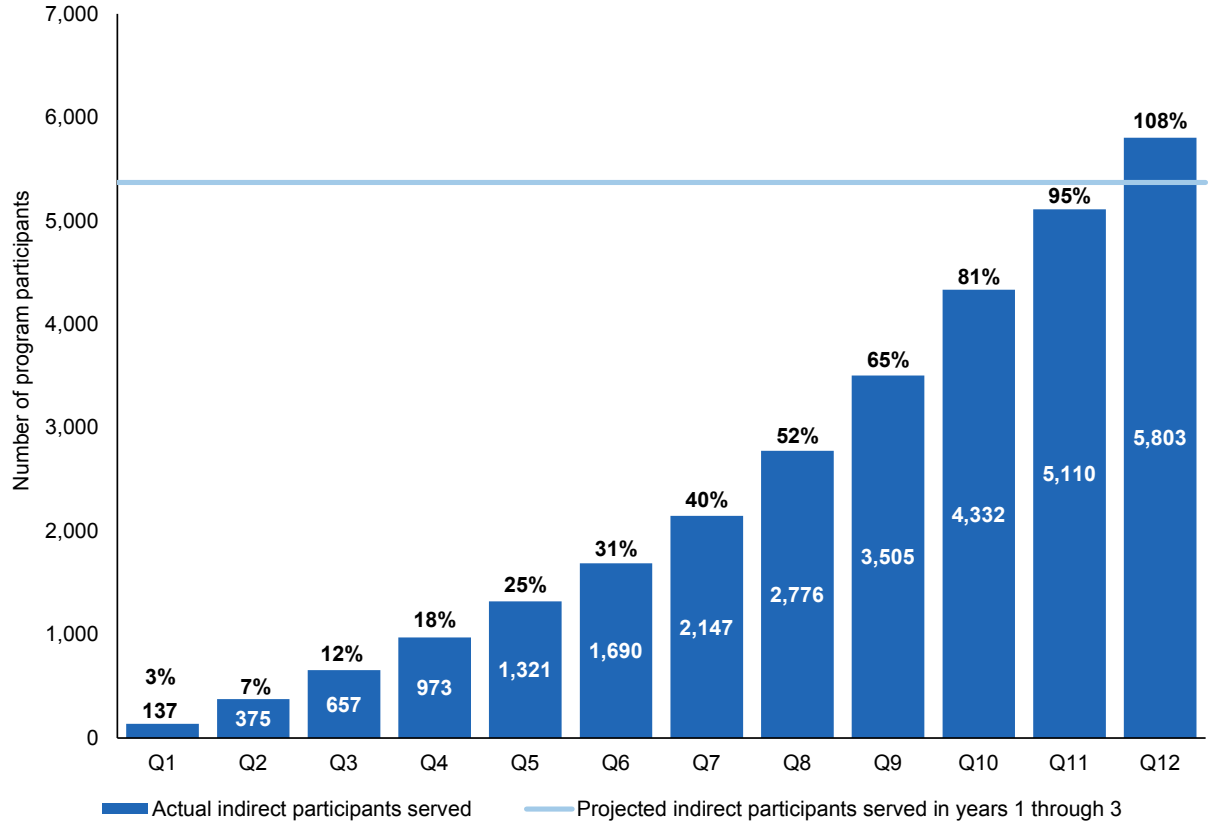
Four Seasons did not make major changes to the definition, identification, recruitment, or enrollment of the target population. Throughout the cooperative agreement, the awardee increased enrollment in many ways, including recruiting new implementing sites, spreading awareness about the program to providers who might be able to refer participants, and improving efficiency in the clinical workflow by, for example, hiring dedicated nurses and documenting best practices.

b. Evidence of enrollment effectiveness

Four Seasons achieved enrollment effectiveness based on its revised enrollment target. Overall, the awardee reported that it enrolled 5,803 participants from September 2014 (when it launched its program) through August 2017, which represented about 108 percent of its 5,371 three-year projected participants (Figure II.1). Four Seasons set this three-year enrollment target in the ninth program quarter, which was a 33 percent reduction from the original enrollment target of 8,000, which it set during the first program quarter. Although Four Seasons met only 73 percent of its original target, all implementing sites steadily increased enrollment throughout the cooperative agreement.

It should be noted that the awardee provided the same palliative care services to all participants who needed it, regardless of their insurance type. However, the enrollment figures reflect only Medicare FFS beneficiaries; participants with other types of insurance are beyond the scope of our evaluation.

Figure II.1. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers. The awardee decreased its original indirect participant enrollment projection from 8,000 to 5,371 at the beginning of Year 3.

c. Barriers and facilitators associated with enrollment effectiveness

The progress Four Seasons made in meeting its three-year enrollment goals was influenced by several factors. High penetration of Medicare Advantage plans in the catchment area, delays in reaching agreements with the implementing sites, and the referring providers' misperceptions about palliative care were barriers to enrollment, whereas outreach and the hiring of new staff facilitated enrollment.

Throughout the cooperative agreement, Four Seasons reported that the biggest barrier to reaching its enrollment target was the high penetration of Medicare Advantage in its catchment area. Awardee leaders originally predicted that twice as many patients would be eligible for the program than were actually eligible because of an unexpected rapid growth in Medicare Advantage plans from the point at which they wrote their application to the start of the cooperative agreement. "I don't know who's been selling [Medicare Advantage plans], but they're very good [at selling the plans]," one site leader said. The HCIA R2 evaluation specified

that only Medicare FFS patients count toward the total number of enrolled participants even though Medicare Advantage patients received the same services as those enrolled in the program. Four Seasons requested a waiver of the FFS requirement, but the request was not approved.

Delays in reaching agreements with the implementing sites also prevented Four Seasons from meeting its enrollment target, according to awardee leaders. These delays were often longer in the larger organizations because the processes for approving new partnerships are complex. Although Four Seasons began recruiting new sites at the start of, or even before, the cooperative agreement, only PCHCV implemented the program and began enrolling participants during the first program year. Two other sites, Greenville Health System and Mission Hospital, did not start enrolling participants until the second program year. A third site, Hospice Care of South Carolina, did not begin enrolling participants until the third program year. In July 2017, the Four Seasons reported that it was close to finalizing an agreement with another site, Frye Regional Medical Center. The awardee planned to help Frye implement the program during the no-cost extension period.

Providers' misperceptions about palliative care undercut the awardee's ability to identify potential participants through provider referrals. All implementing sites said that it was difficult to encourage providers to refer participants to the program, especially during the early stages of implementation. The referrals were perhaps the most important mechanism for identifying potential participants because data from claims and electronic medical records (EMRs) were usually not sufficient. By the third program year, the awardee reported that all sites had had successfully engaged referring providers and provider organizations through education and partnerships. Their strategies for doing to included (1) holding annual education events for providers at hospitals about hospice and palliative care, (2) establishing liaisons with hospital providers in order to increase communication and collaboration, and building relationships with outpatient clinics to recruit participants in this setting.

In addition to outreach efforts to referring providers, two implementing sites in particular facilitated enrollment by hiring staff to help identify and potential participants. The two sites each hired a full-time nurse whose positions were supported by funding from Four Seasons and the cooperative agreement. The nurses coordinated program activities at their sites in order to make the program operate more efficiently and to reduce the burden of enrollment and data collection on program staff.

2. Delivery of program services

a. Description of and changes to service delivery model

Before the cooperative agreement, Four Seasons operated the CPC program at its own site. The awardee used HCIA R2 funding to standardize and document aspects of its program operations, such as enrollment, data collection, participant assessments, service delivery, and clinical workflow. The standardization and documentation allowed Four Seasons to help new organizations implement the CPC program.

The health care services provided by the CPC program did not change throughout the cooperative agreement. The awardee sought to address participants' needs holistically by, for example, providing spiritual and social support as well as clinical care. The highly collaborative,

> "Adding clinics can double the number of patients we serve. In the clinics, there is lower overhead and greater productivity because [providers] don't have to travel and can document in real time, which is harder when providers travel to see patients."
>
> —Site leader

multidisciplinary CPC care teams integrated inpatient and outpatient care such that it spanned all settings through which participants with advanced illnesses transition, such as hospitals, clinics, private residences, nursing homes, and assisted living facilities. The CPC care teams were headed by a nurse practitioner (NP) or a physician's assistant (PA) who oversaw registered nurses (RNs), social workers, and administrative support staff. The care provided by the teams focused on symptom management, quality of life, psychosocial supports, coordination with community-based resources and other health care providers, advance care planning, shared decision making, and spiritual support. In the ninth program quarter, Four Seasons began enrolling participants in a new tele-health model, also partially funded by the cooperative agreement. The model allowed providers to offer the palliative care services delivered on site to participants in rural counties via telephone or videophone. During the third program year, two implementing sites started multiple palliative care outpatient clinics in an effort to enroll more outpatient participants and increase access to palliative care services. One site leader described outpatient clinics as "cost-effective and efficient" for the implementing sites when compared with other inpatient or home-based services.

CPC protocols gave staff the flexibility to tailor services to the needs of an individual participant but provided guidance on the timing of encounters. Starting in the third program year, the interview respondents reported that they consistently assessed participants' risk levels by using a risk assessment tool that classified participants into high and low acuity levels. According to the protocols, CPC care teams were to schedule in-person home appointments within 48 hours of enrollment for high-risk participants; for low-risk participants, the care teams were to schedule a home visit within 7 to 10 days and an in-person clinic visit within two weeks of enrollment. Care teams followed up with participants in person or by phone as needed throughout the remainder of their enrollment. During the first encounter, program staff typically assessed the participant's health, developed a care plan with input from the participant and caregivers, and documented decisions for advance planning. Other services included symptom management, social work, education in disease management, support with complex medical decisions, psychosocial support, and participant and family education about palliative care.

Four Seasons originally used a centralized administrative team and nursing staff to support its sites. In the third program year, however, the awardee established regional leaders, who offered guidance on the best way to tailor the program to providers in their areas. This approach resulted in some regional differences in service delivery. Variation in provider caseloads is a good example of these differences. The optimal caseload for NPs and PAs in urban areas differed from the optimal caseload in rural areas because in the latter, providers spent more time traveling to participants (80 versus 60 participants per clinician, respectively). RNs generally had a caseload of 150 to 200 participants. Awardee leaders said that caseloads were purposely kept low in order to allow clinicians to provide "high-touch" care to participants, who generally had complex conditions and multiple comorbidities.

Four Seasons and its implementing sites used EMR and other data systems to track the status of participants. The reports generated from the tracking systems were reviewed weekly by an

interdisciplinary group composed of a clinician, a chaplain, and a social worker who discussed high-acuity participants and recommended interventions that the CPC care teams could use to manage the participants' care.

Program staff were required to attend a 40-hour immersion course soon after they were hired. Four Seasons hosted the immersion course multiple times a year based on demand. The course spanned numerous topics and activities including but not limited to (1) classroom lectures about palliative care, cultural competency, and other relevant topics; (2) role-playing of complex cases through which program staff learned how to navigate difficult conversations and family dynamics; (3) best practices for coding and billing; and (4) workplace leadership. Non-clinical program staff also attended the immersion course to better understand the program's values and services so that they could better support clinical staff.

HCIA R2 funding also supported the program's activities to educate program staff, program clinicians, external providers, participants, and families/caregivers about palliative care, and these efforts included online educational modules on palliative care. The awardee also gave presentations at hospitals, nursing homes, physicians' offices, and assisted living facilities to educate clinicians about palliative care. During the presentations, the awardee also distributed informational materials and encouraged attendees to refer participants to the CPC program. During the third program year, Four Seasons developed and posted to its website educational modules about palliative care targeted to participants, their families, and providers. The modules covered topics such as advance care planning, coping with pain and other end-of-life issues by using mindfulness and meditation techniques, coping with dementia, and treating and coping with chronic obstructive pulmonary disease. The awardee primarily promoted the educational modules on its website to CPC participants and referring providers but reported that it eventually planned to publicize the materials nationwide after updating the website with more information and resources.

b. Evidence of service delivery effectiveness

Overall, the awardee successfully achieved service delivery effectiveness. We could not, however, fully assess this because we did not have awardee-reported service delivery measures. Nonetheless, the results from surveys and interviews indicated that program staff were able to deliver program services and engage participants as intended. The awardee was also able to hire and retain needed program staff throughout the cooperative agreement despite the challenges in the palliative care field with workforce shortages and staff burnout. Four Seasons also successfully trained program staff and engaged providers and provider organizations in referring participants.

Delivery of intervention services. Interview and survey respondents indicated that the program services were delivered as intended, although the awardee did not report data on the timeliness of participant encounters or on the average duration of participant enrollment. Ninety and 50 percent of program staff and clinicians, respectively, who answered the survey question on the program's impact on their ability to respond to participants' needs in a timely way believed that the program had a positive impact; 10 and 45 percent of staff and clinicians reported that the program had no impact; and less than 5 percent in each group believed that the

program had a negative impact. Moreover, the program protocols did not prescribe a duration for receiving program services.

Throughout the duration of the program and across all sites, 80 percent or more of program services were delivered in person. Higher percentages of in-person encounters took place during the first two program years before the tele-health pilot was launched. The remaining 20 percent or less of encounters occurred via the tele-health program, a telephone visit, or e-mail or online counseling.

Program staff and leaders at multiple implementing sites throughout the duration of the cooperative agreement reported that the program's protocols allowed the sites to successfully adapt the program to their own setting and region. As discussed, providers in rural areas carried smaller caseloads than their counterparts in urban areas because of longer travel times in rural areas. As another example of the flexibility of the program, one site implemented a home-based palliative care program after finding that palliative care participants discharged from the inpatient setting had trouble making appointments for outpatient palliative care.

Engagement of program participants. The program leaders we interviewed as well as 70 percent of the program staff we surveyed believed that they successfully engaged participants. The program's discharge rate of 37 percent may seem high, but according to awardee leaders, this rate is consistent with the discharge rate in other large palliative care programs. Awardee leaders explained that the discharge rate captured participants whose goals of care had been met, who had improved in their health status, or who chose to stop services. Many participants who had been discharged may have re-enrolled if their conditions worsened, according to awardee leaders. "We know at least 20 percent [of discharged participants] come back within that year, and 40 percent are [transitioned to] hospice." Given that people with life-limiting, advanced illnesses are the target population, an awardee leader estimated that the majority of the remaining discharged participants probably die.

Four Seasons did not report the number of patients who declined palliative care services. According to the survey results, program staff and clinicians believed that the most common reason for refusing program services was that patients did not understand the program and/or palliative care itself. Moreover, awardee leaders believed that a very small percentage of discharged participants chose to leave because they did not want to pay co-payments for program services. The second most common reason for declining services was, according to our survey results, patients' concerns about insurance coverage or the financial burden of the program. To overcome this challenge, awardee leaders said that they made it a point to provide palliative care services to patients regardless of whether they paid their co-payments or chose to be enrolled in the HCIA R2–funded program.

Four Seasons designed multiple methods for evaluating participant, family, and caregiver experience and satisfaction with program services but was unable to report the final results of these efforts when we wrote this report. However, awardee leaders did report preliminary results from the telephone survey of participants' and caregivers' satisfaction administered by Four Seasons. The results showed that overall satisfaction with program services was above 80 percent, which awardee leaders saw as "a high number within health care." One limitation of the survey was that not all implementing sites could administer it because participants at some sites

were not able to distinguish their experience with program services from other types of care received from the hospital or health system. Four Seasons also partnered with two research teams at the University of Massachusetts and Duke University to study the experience of families, caregivers, and tele-health participants. Data collection for these studies had either not begun or was not completed when we wrote this report.

Four Seasons leaders believed that using risk assessments to determine the intensity of follow-ups improved participant satisfaction and reduced the number of participants who dropped out of the program. "In the beginning, we probably lost some [participants] because they weren't seen frequently enough, or they didn't feel that we were making a difference. But now, I think it's definitely better," an awardee leader said.

Four Seasons leaders also estimated that they provided tele-health equipment to one-third of the tele-health participants because these participants did not have personal access to the technology required to participate in the tele-health model (smartphone, computer tablet, or access to the Internet). Awardee leaders reported that the tele-health technology had not been a barrier to engaging tele-health participants, even for older patients, who, according to the awardee leaders, often have difficulty using new technology. "Fifteen percent of our patients who are using tele-health are over 90 years old, and amazing[ly], we can get compliance with [older patients] even on a technology-based system."

Staffing and training. Four Seasons was partially successful in hiring program staff during the cooperative agreement. Awardee leaders and staff reported that hiring qualified clinicians during the beginning of the cooperative agreement was a challenge, but insufficient staffing did not seem to negatively affect the awardee's ability to deliver program services, according to the interviews that we conducted in the third program year. Four Seasons reached 75 percent of its overall target for new-hire, full-time employees (FTE)—13.5 of 18 FTE. The awardee was closer to reaching its annual new-hire FTE target during the first and third program years (91 and 100 percent, respectively), but it reached only 37.5 percent of the target in the second program year.

Interview respondents did not perceive that failing to reach the staffing target negatively affected the awardee's ability to deliver program services as intended, although some respondents indicated that insufficient staffing may have contributed to staff and clinician burnout throughout the cooperative agreement. Overall, most program staff and clinicians we surveyed did not find their workload or caseload to be too heavy (78 and 71 percent, respectively). However, 50 percent of clinicians reported that the program increased their feelings of burnout a little or a lot. Program leaders at the Four Seasons site began monitoring staff and clinician burnout at the start of the cooperative agreement, saying that "staff burnout is always a concern in this line of work [palliative care]." Four Seasons monitored burnout through the TeamSTEPPS Teamwork Perceptions Questionnaire by the Agency for Healthcare Research and Quality;[6] the awardee also tried to prevent burnout by managing the workload and schedules of clinical staff. Other implementing sites we interviewed did not mention anything about using specific strategies to prevent burnout, although one site leader said that palliative care clinicians are accustomed to burnout and therefore can cope with it. Clinical staff at one site stated that

[6] See https://www.ahrq.gov/teamstepps/instructor/reference/teamperceptionsmanual.html.

burnout was a bigger problem during the beginning of the cooperative agreement, but that eventually, the committed, strong, passionate team of people hired by the awardee helped to reduce the burden on individual team members.

During the beginning of the cooperative agreement, Four Seasons also reported that it was concerned that clinician turnover would negatively affect its ability to provide program services. Multiple interview respondents perceived that turnover in their regions was high because providers prefer not to live in largely rural western North Carolina and South Carolina, and the salary for those who do provide palliative care services is relative low. In the third program year, however, the awardee no longer reported turnover as a concern. According to interview respondents, new hires during the third program year were a result of (1) normal turnover, (2) replacing staff who were let go because of performance issues, and (3) the need to meet a sudden increase in the demand for palliative care services in western North Carolina after a large hospital in the region closed its palliative care program.

Finally, Four Seasons successfully trained program staff and leaders. Multiple people in both roles reported that training for the program, and especially the immersion course, continued to be a valuable opportunity throughout the cooperative agreement. Of the 77 percent of clinicians and the 80 percent of staff who had completed the training by the time we administered the survey, 82 and 90 percent of these respondents, respectively, indicated that the training helped to improve their performance in the CPC program.

Recruitment and engagement of providers. Despite an initial delay, the awardee was ultimately successful in recruiting four provider organizations to implement the program (see also Section II.B.1.c). Delays in finalizing partnerships with these implementing sites affected the awardee's ability to meet its enrollment goals.

Moreover, once the program was implemented, Four Seasons successfully engaged the four provider organizations. All interview respondents said that the Four Seasons central office staff adequately engaged leaders, clinicians, and staff at the implementing sites by being available one-on-one by phone whenever necessary in addition to holding formal monthly meetings to discuss implementation challenges, specific participants, and data on progress and performance.

c. Barriers and facilitators associated with service delivery effectiveness

The service delivery effectiveness of the CPC program was influenced by several factors. The barriers included limitations in health information technology (health IT) system, relatively low reimbursement for palliative care services, and disincentives for accountable care organizations (ACOs) to participate in the program. However, training of program staff, creating regional managers, and overcoming misperceptions about palliative care through outreach and education efforts to providers and the community reportedly helped the program staff and clinicians to deliver the program services as intended. According to Four Seasons, however, several factors helped program staff and clinicians to deliver program services as intended. These facilitators included the training of program staff, the use of regional managers, and the use outreach and education to overcome the misperceptions that providers and communities have about palliative care.

Throughout the cooperative agreement, Four Seasons reported that it was a challenge to use the Quality Data Collection Tool (QDACT) to collect, analyze, and present program data to make quality improvements in the CPC program and engage stakeholders. The awardee tried but was unable to find an adequate replacement for QDACT during the cooperative agreement. The challenge that Four Seasons reported most often was that QDACT required clinical staff to double enter data, first into the EMR and then into QDACT. Awardee leaders believed that this responsibility in particular contributed to the clinical staff's burnout during the first program year, and it continued to burden them throughout the cooperative agreement.

At least two implementing sites hired a CPC nurse to manage QDACT activities and lessen the burden on other clinical staff, but one respondent described this as a temporary "bandage" solution: "For the long term, it is not enough. The program will not be sustainable until QDACT is integrated into the EMR." Moreover, one program leader noted that QDACT was limited in its ability to provide more sophisticated data, such as information on participant outcomes. "We could have used that data to advocate for more staffing or for expanding the palliative care program," the respondent explained. On the other hand, leaders at two implementing sites reported that the QDACT data helped them to improve palliative care services. For example, one site leader said that process measures related to non-medical issues helped clinicians not trained in palliative care to focus on a participant's holistic well-being.

In addition to QDACT, Four Seasons said that finding an EMR system that aligned with the requirements of delivering palliative care services was one of its biggest challenges throughout the cooperative agreement. The lack of a suitable EMR made it difficult for providers to efficiently provide such services as evaluating participant outcomes, communicating with other providers and participants, and tracking and monitoring participants, especially those who had transitioned from an inpatient to an outpatient setting. Awardee leaders reported that existing EMR systems that support palliative care are not certified, which will result in penalties starting in 2018 under the Quality Payment Program's Merit-based Incentive Payment System. Sixty-eight percent of the program staff we surveyed reported that EMR and other technology were barriers to achieving program goals.

> "The lesson learned is [that] until we have an electronic health record that works for patients, families, and clinicians across multiple settings, it's going to be really hard to deliver good, consistent palliative care."
>
> —Awardee leader

Program leaders at Four Seasons and at the implementing sites reported that it was difficult to make palliative care financially viable throughout the cooperative agreement. They noted that although palliative care services can save money for payers such as Medicare, the reimbursement rates in their communities for providers are so low that the clinical organizations ultimately lose money on palliative care services. The sites informed us that provider organizations view palliative care as a loss leader for generating the more financially viable hospice revenue. However, one awardee leader explained that the organization could not increase its transition rate from palliative care to hospice without hiring more clinicians and support staff, which would require up-front funding. In order to offset the low reimbursement rates for palliative care, program leaders used different strategies, like supplementing the care team with less expensive staff such as social workers and nurse navigators, and by balancing more efficient clinic visits with more expensive home-based visits.

Four Seasons leaders posited that the structure of incentives at ACOs may have initially made the implementing sites that are members of an ACO uncertain of the benefits of implementing the program. "ACOs have been kind of late to endorse and embrace palliative care [because they are already] doing things in the community, like hiring caseworkers [for managing] patients with chronic illnesses," one leader explained. The ACOs' reluctance can be exacerbated by their concerns about how attribution works for participants receiving palliative care services through the program. That is, if an implementing site delivered the majority of the services, then the ACO could not count the patient in its population. Even after finalizing agreements, site leaders said that implementing the program took time, as each site determined how to adapt the program and find adequate resources such as staff.

Despite these challenges, Four Seasons continued to receive positive feedback about the training it provided to program staff and clinicians. For instance, interview respondents continued to applaud the immersion course for superb training on community-based palliative care. One site leader said that the immersion course was valuable enough to send clinicians to participate, despite the huge cost to the organization (clinicians are not billable during the five-day immersion course). In addition to the immersion course, Four Seasons central office staff provided ongoing training to implementation sites through weekly or monthly calls, summary of data from QDACT and other sources, and one-on-one communication with members of the team as needed. One respondent described the ongoing training from Four Seasons as "solid and continuous" support that helped the implementing sites to "focus on what we need to improve."

> "Through our weekly or monthly calls, we have maintained a really good relationship [with implementing sites]. They've begun to understand the importance of palliative care—not only inpatient but also outpatient palliative care."
>
> —Awardee leader

Awardee leaders believed that decentralizing program management to the regional managers made the program more efficient and sustainable by allowing service delivery to be adapted by people who know and understand the unique challenges in their regions. For example, awardee leaders reported that there can be important differences between the more populous Henderson County and the mountainous or rural areas in such factors such as patient populations, efficient billing practices, and preferred encounter type or workflow. There may be similar differences depending on the setting, such as a large hospital system versus a small clinic.

Misperceptions about palliative care among participants and their families have always challenged both enrollment and service delivery, according to Four Seasons staff. Interview respondents, however, reported that the efforts of Four Seasons and its implementing partners to educate the community have been successful. The awardee has been working on expanding its community outreach efforts by launching online patient and family education modules on its website.

C. Assessment of perceived program effects on the delivery of care and outcomes

According to all interview respondents and almost all survey respondents, the program achieved its intended outcomes goals. They attributed this success to the program's positive effect on participant satisfaction and quality of life, and to the fact that it reduced spending and

utilization. Leaders at the implementing sites reported that internal evaluations of the program's impact also showed reduced spending and utilization. For example, one implementing site reported that "large improvements" in dyspnea and in the appropriate use of oral medications resulted from effective medication reconciliations for the majority of patients. The site also found that the program saved money. Another implementing site had not calculated the exact cost savings at the time of our interviews but said that its analysis showed a reduction in other utilization outcomes that are linked to costs. For example, there was a 19 percent reduction in readmissions for participants receiving home health services.

Moreover, almost all program staff and clinicians surveyed believed that the program achieved its goals, although a small number of program staff surveyed said that it was too soon to tell. Most program staff and clinicians believed that the program had a positive impact on participant satisfaction, although more than one-third of clinicians believed it was too soon to tell. Nearly all program staff and clinicians who responded to the question about whether the program made a difference in terms of meeting critical needs of the community believed that it did so. Finally, nearly all program staff and 59 percent of clinicians surveyed believed that the program had a positive impact on the providers' ability to help participants achieve their health goals.

Respondents also believed that program services had a positive effect on the satisfaction of participants, caregivers, and families because they helped participants to manage pain and other symptoms through patient-centered care. CPC providers focus on "listening to [participants] and honoring their wishes," explained one program leader. Another interview respondent felt that advance care planning in particular improved the participants' quality of life by empowering them to set and achieve their own goals.

Leaders and staff at all implementing sites believed that CPC saved money by intervening early in the continuum of care, thereby increasing the number of people who use hospice and reducing expensive treatments of questionable benefit during inpatient stays. "[Hospice] is one of the highest savings for Medicare health care costs," one respondent said. Another respondent explained that palliative care helps patients decide which treatments are the best way to meet their personal goals (palliative care does not generally include the intensive care unit, high-tech medical care, or other expensive care). Moreover, interview respondents believed that research shows that palliative care saves money and improves the quality of life, particularly for high-cost Medicare beneficiaries receiving home-based care.

Awardee leaders also believed that the program's telemedicine services improved participants' access to palliative care services by expanding the program's catchment area, particularly to rural counties in which access to palliative care services was minimal. These services not only "kept people out of the hospital [by] catching symptoms early on and responding to them [quickly]" but also improved participants adherence to medications and self-care regimens.

D. Implications of implementation findings for the design and interpretation of an impact analysis

Interview and survey respondents overwhelmingly expressed their belief that program services delivered throughout the cooperative agreement will have an impact on the outcomes that they are intended to affect. Respondents said that (1) program services were delivered as intended and (2) the program's palliative care services are valuable because they encourage providers to take more time to talk and listen to participants about their goals and health status than other providers would. However, because we did not have awardee-reported data on, or targets for, service delivery, it was difficult to assess whether enough program services were provided to affect participant outcomes.

Despite their belief that the program will achieve its intended outcomes, respondents also noted that the following factors may make it difficult for an impact evaluation to detect the effects that the program was expected to achieve:

- **Low enrollment during the first two program years.** Low enrollment and high penetration of Medicare Advantage may make it difficult to detect program impacts. Four Seasons met 73 percent of its three-year enrollment target, although 63 percent of the total 5,803 participants were enrolled during or after the eighth program quarter; these participants may have received less exposure to the program. Moreover, the awardee expected to enroll more Medicare FFS beneficiaries than it actually did because of a rapid and unanticipated increase in Medicare Advantage beneficiaries in its catchment area.
- **Suboptimal participant mix.** Improvements in participant outcomes will likely be driven by participants who are either seriously ill or making the transition from one type of care or another, according to one site leader. This respondent also expressed concern that these patients may be under-represented in the enrolled participant population. Moreover, the impact evaluation may not pick up outcomes of participants who were seen in outpatient palliative care clinics because most of the clinics began to operate in the third program year. Therefore, the participants' exposure to program services will not have been long enough to produce a detectable effect.
- **Misaligned payment incentives.** Finally, one respondent believed that health systems incorporate incentives that encourage physicians to send high-cost patients to hospice rather than palliative care. The rationale is that, once these patients are in hospice, they may be "carved out of" their population health average per-patient costs. Because of this payment incentive, CPC programs may not be able to rely on these patients as a revenue source.
- **Inadequate staffing, especially during the first two program years.** As discussed, workforce shortages made it difficult to hire experienced staff, especially during the first two program years. The immersion course eventually helped Four Seasons to train clinicians who were not trained in palliative care, according to one site leader.
- **The awardee's limited ability to use program data to make qualitative improvements in program implementation.** Although the cooperative agreement made it possible for the implementing sites to collect more program data, awardee leaders said that the lack of access to claims data in addition to the inadequacies of QDACT made it difficult to be "nimble and flexible at making changes [given] results of that data."

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Four Seasons Compassion for Life's CPC program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Four Seasons

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) population with 6 months of program exposure by February 28, 2018	4,779[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	704
Likelihood of all-cause hospitalizations	314
MDE sample size requirement to detect 20% effect	
Total expenditures	176
Likelihood of all-cause hospitalizations	79
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Intervention implemented that differs from baseline period	No, low rate of identification of treatment group with claims data with significant dilution of treatment effect within an intent-to-treat framework
Likelihood of solid comparison group	Serious concern; we may be not able to identify a strong comparison group
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Lack of strong comparison group
Survey data for treatment group that will be analyzed	Clinician and staff surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

Although we do plan to carry out an analysis that compares outcomes of program participants with those of a selected comparison group, we are not prepared to characterize that analysis as a rigorous assessment of the impact of the CPC program. Enrollment in the program is based partly on subjective physician judgment and considers traits such as housing status, lack of caregiver support, and most important, the referring provider's assessment of the likelihood of death in the next year. These traits cannot be identified in Medicare claims or other data. Under such circumstances, our confidence in a comparison group drawn solely on the basis of information available in Medicare claims and enrollment data is naturally limited.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents our third summary of the beneficiary characteristics, common claims-based outcomes, and awardee-specific claims-based outcomes at baseline for beneficiaries participating in the CPC program. Based on the awardee's eligibility criteria, the treatment group being assessed in this report consists of Medicare FFS beneficiaries age 65 or older who had a life-limiting disease and enrolled in the program.

As of May 31, 2016, the CPC program had 2,116 participants.[7] The enrollment date is defined as the date that a beneficiary began to receive palliative care from Four Seasons, which may be earlier than the date that the beneficiary received the letter of implied consent regarding participation in the HCIA R2 program.

In presenting baseline characteristics for this report, we restricted the treatment group to beneficiaries (1) who enrolled in the CPC program from September 2014 through May 2016, (2) who were enrolled in Medicare FFS for at least 90 days during the baseline year (the 365 days immediately before their enrollment), (3) who were enrolled in Medicare FFS in the month that they enrolled in the CPC program, and (4) for whom we were able to link Medicare claims and enrollment data. After we excluded beneficiaries who did not meet these criteria, a total of 2,050 participants were included in the analysis of baseline characteristics for this report. The calendar period covered by the baseline quarters is based on the enrollment date for each beneficiary and therefore varies by beneficiary.

A major challenge for the impact analysis is replicating the program's auxiliary eligibility criteria, which are outlined in a draft tool for providers to use when considering whether to refer a patient to the program. Some of the auxiliary eligibility criteria are relatively easy to apply via claims data, such as a high use of resources in the period before a beneficiary enrolled in the program (frequent hospitalizations and ED visits). Other criteria are much more challenging to apply via claims data, such as end-stage dementia, physical limitations, polypharmacy, a

[7] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to Mathematica. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to Mathematica before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

palliative performance scale, and the risk of falling. We have tried to determine which beneficiaries would have been eligible for the program had it existed in the pre-intervention period (and which beneficiaries are eligible for the comparison group) based on the auxiliary criteria, because these criteria are what the awardee uses when selecting participants. To try to make this determination, we have used Medicare enrollment and claims data to characterize beneficiaries in the CPC program, emphasizing the measures that are included in the awardee's tool for referral to the program. Although our goal has been to identify study eligibility criteria that we could use to select a comparison group (such as mortality, diagnoses, prior utilization and spending, in combination with demographic characteristics, HCC score, and dual-eligible status), the analyses we have conducted suggest that this may not be possible.

In terms of demographic characteristics, the Medicare FFS beneficiaries in the CPC program are much older and sicker than the general Medicare population in the Four Seasons service area, which is not surprising given that all participants have a life-limiting illness with a prognosis of three or fewer years. Close to half (43 percent) of the CPC participants are 85 years or older and about a third (34 percent) are 75 to 84 years old (Table III.2). In comparison, the average age of Medicare beneficiaries age 65 and older in the Asheville, North Carolina, Hospital Referral Region (HRR)—the HRR in which most CPC participants live—is 75 years old.[8] Most CPC participants (85 percent) were originally entitled to Medicare because of age; almost all of the other beneficiaries were originally entitled to Medicare because of a disability. The majority of participants are female (60 percent). Nearly all of the participants are white (95 percent), which reflects the racial composition of the Four Seasons service area. In 2014, only 2 percent of Medicare beneficiaries age 65 or older in the Asheville HRR were black. One-quarter of participants are eligible for both Medicare and Medicaid. The average hierarchical condition categories (HCC) risk score of 3.46 for CPC participants is nearly four times as large as the average HCC risk score of 0.87 for Medicare FFS beneficiaries age 65 or older in the Asheville HRR—indicating substantially poorer health status and greater needs for care among the CPC participants.

Consistent with this poor health status, CPC participants had high rates of Medicare expenditures and service use in the year prior to enrollment, particularly in the quarter immediately before enrollment. In Table III.3, we report baseline utilization and expenditure data for a common set of measures, including the four core measures from the Center for Medicare & Medicaid Innovation. By providing a continuum of palliative care services that addresses participants' needs and by integrating care across settings, the awardee expects to reduce expenditures by reducing hospitalizations and ED visits compared with utilization that would have taken place absent the CPC program. We examined baseline expenditures by calculating average per beneficiary per month (PBPM) Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $3,026, with average PBPM Medicare payments for inpatient ($1,137) and skilled nursing facility ($514) services being the largest drivers of total cost of care. Although Four Seasons does not require acute care utilization in the baseline period to be eligible for the program, providers consider the

[8] The data for the Asheville HRR presented here and in the following paragraphs are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation; HRR Table—Beneficiaries 65 and Older." Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed April 2016.

number of recent hospitalizations and ED visits when determining whether a patient is appropriate for the program. Therefore, it is not surprising that the total average PBPM Medicare payment in the last quarter before enrollment ($5,852) was more than twice as large as in any other quarter.

As with expenditures, CPC participants on average had high use of expensive Medicare services before enrollment in the program, particularly in the last quarter before enrollment. The annual rate of acute care hospitalizations was 1,390 per 1,000 CPC participants during the baseline period—much higher than the rate of 215 per 1,000 Medicare beneficiaries age 65 or older in the Asheville HRR in 2014. The annual rate of acute care hospitalizations was 3,292 per 1,000 CPC participants during the last quarter before enrollment. This figure is higher than the rate calculated from the February 2016 finder file, which was an annual rate of 2,170 per 1,000 CPC participants during the last quarter before enrollment. This change likely reflects an increase in the share of CPC participants admitted to the program during or immediately after a hospitalization because the program expanded its referral sources to include additional hospitals. The likelihood of a 30-day unplanned readmission for CPC participants was 18 percent per discharge or 10 percent per beneficiary. The annual rate of ED visits not leading to a hospitalization was 1,183 per 1,000 CPC participants, while the rate of observation stays was 171 per 1,000 CPC participants. Thus, there may be an opportunity to reduce potentially avoidable admissions, ED visits, observation stays, and readmissions through comprehensive palliative care. At baseline, CPC participants were very high users of primary care visits (10,768 primary care visits in ambulatory settings per 1,000 CPC participants per year) and specialist visits (9,529 specialist visits in ambulatory settings per 1,000 CPC participants per year).

We also examined two other measures of utilization in the baseline period: (1) hospice and (2) intensive care unit or coronary care unit (ICU/CCU) services. Over the baseline period, the likelihood of hospice use for CPC participants was 5 percent, which was larger than the likelihood of 3 percent for all Medicare beneficiaries age 65 or older in the Asheville HRR in 2014 (Table III.3). The likelihood of ICU/CCU use for CPC participants in the baseline period was 15 percent. That rate was driven by much higher ICU/CCU use in the last quarter before enrollment (likelihood of 9 percent) than in each of the prior three quarters (likelihood of 2 percent to 3 percent).

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Four Seasons' program through May 31, 2016

	All participants (N = 2,050)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	1	0.05
65 to 74	471	23
75 to 84	705	34
85 and older	873	43
Gender		
Female	1,225	60
Male	825	40
Race		
White	1,943	95
Black	81	4
American Indian, Alaska Native, Asian/Pacific Island American, or other	18	0.88
Hispanic	5	0.24
Original reason for Medicare eligibility		
Old age and survivor's insurance	1,736	85
Disability insurance benefits	305	15
End-stage renal disease (ESRD)[a]	9	0.44
Hospice[b]	63	3
Medicare/Medicaid dual status, percent dual[b]	522	25
HCC score[c]		Statistic
Mean		3.46
25th percentile		1.82
Median		3.16
75th percentile		4.65

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the beneficiary began to receive palliative care from the awardee. All beneficiary characteristics are measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition categories.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Four Seasons' program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	2,050	2,030	2,033	2,047	2,050
Average Medicare expenditures PBPM[a]					
Total	3,026 (66)	1,914 (90)	2,025 (89)	2,286 (92)	5,852 (143)
Acute inpatient	1,137 (34)	515 (39)	546 (44)	625 (47)	2,845 (98)
Inpatient other[b]	129 (15)	84 (35)	80 (20)	97 (24)	254 (32)
Outpatient[c]	450 (20)	378 (23)	395 (24)	457 (27)	567 (28)
Physician services	469 (11)	329 (14)	343 (15)	388 (14)	812 (20)
Home health	166 (6)	128 (9)	145 (9)	162 (10)	227 (11)
Skilled nursing facility	514 (20)	335 (30)	356 (32)	383 (31)	978 (44)
Hospice	107 (12)	90 (13)	106 (15)	122 (15)	108 (12)
Durable medical equipment	56 (6)	54 (7)	54 (6)	53 (6)	62 (8)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	1,390 (33)	689 (41)	705 (44)	852 (47)	3,292 (78)
Outpatient ED visits	1,183 (40)	865 (51)	922 (57)	1,110 (59)	1,826 (79)
Observation stays	171 (10)	156 (18)	128 (16)	151 (18)	248 (22)
Primary care visits in any setting	15,706 (301)	11,151 (350)	11,950 (368)	13,514 (378)	26,088 (569)
Primary care visits in ambulatory settings	10,768 (204)	8,703 (233)	9,234 (246)	10,304 (261)	14,773 (329)
Specialist visits in any setting	15,873 (349)	11,594 (378)	12,238 (392)	13,767 (437)	25,773 (691)
Specialist visits in ambulatory settings	9,529 (225)	8,475 (251)	8,770 (254)	9,579 (273)	11,252 (282)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	70 (1)	14 (1)	14 (1)	17 (1)	58 (1)
Percentage with an outpatient ED visit[e]	54 (1)	16 (1)	16 (1)	20 (1)	30 (1)
Percentage with an observation stay[f]	14 (1)	4 (< 0.5)	3 (< 0.5)	4 (< 0.5)	6 (1)
Percentage with a 30-day readmission among all discharges	18 (1)	13 (2)	16 (2)	19 (2)	21 (2)
Percentage of participants with a readmission among all participants	10 (1)	2 (< 0.5)	2 (< 0.5)	3 (< 0.5)	6 (1)
Percentage of participants who used hospice	5 (< 0.5)	2 (< 0.5)	0 (< 0.5)	1 (< 0.5)	1 (< 0.5)
Percentage of participants who used ICU/CCU	15 (1)	3 (< 0.5)	3 (< 0.5)	3 (< 0.5)	9 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; ICU/CCU = intensive care unit/coronary care unit; PBPM = per beneficiary per month.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Four Seasons has partnered with the American Academy of Hospice and Palliative Medicine (AAHPM) to develop a hospice- and home care–based alternative bundled payment model. The model is based on data and experience that the awardee gained from the CPC program, but a lag in claims data delayed its development and implementation. Four Seasons will continue to develop and refine the model by working with other palliative care organizations and national leaders.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Four Seasons has continued to develop its payment model on a PBPM basis for the CPC program. The main purpose of the model is to cover a standard set of services on a monthly basis, including advance care planning, up to three goals-of-care conferences, home visits, clinic visits, symptom management, coordination of services, social work, and a mix of services provided by a hospice team. Services not related to palliative care—including hospitalizations, primary care, and specialty care—are carved out of the payment model.

C. Status of the payment model

At the end of the third program year, Four Seasons was still working with researchers from Duke University to collect data, conduct an internal evaluation, and develop a payment model for the CPC program. During the second program year, Duke conducted a time, motion, and cost study to capture the time, resources, and costs required to deliver program services. The awardee originally planned to implement the payment model during the third program year based on data collected during the first two program years. Because of delays in data collection, Four Seasons modified the plan to perform a simulation of the payment model during the third program year by using claims data to show how the money would have flowed. However, an additional delay in receiving needed data interfered with this plan as well. Four Seasons had recently received claims data at the time of our interview and informed us that it would start to compute the magnitude of a palliative care services bundled payment that it helped AAHPM to design based on a comparison between the actual payments for similar patients who received conventional care and the simulated payments based on the proposed bundled payment. Four Seasons will continue this analysis after the end of the cooperative agreement (including during the no-cost extension), and it is working with Duke to raise funds for this effort.

As mentioned above, Four Seasons has played an important role in the AAHPM's Alternative Payment Model Task Force and has worked with other national palliative care leaders to develop a bundled payment model on a PBPM basis. In particular, the awardee used

data collected from its HCIA R2 program to help design the payment model framework and to provide evidence that supports the use of intensive palliative care in the target patient population.

The purpose of the AAHPM proposed payment model is twofold. First, it is expected to provide a payment structure that can be available to the broadest possible set of community-based palliative care providers. Second, it is expected to maximize access for Medicare beneficiaries to high-value palliative care services. The Patient and Caregiver Support for Serious Illness (PACSSI) alternative payment model proposed by AAHPM is intended to address several limitations of the current Medicare payment schedule by doing the following: (1) making payments for palliative care and support services available to non-billing clinicians such as nurses, social workers, pharmacists, or spiritual care professionals; (2) filling in the gaps left by FFS payments for chronic care management, complex chronic care management, and remotely-delivered services to support high-quality palliative care; and (3) allowing palliative care teams to participate as alternative payment model entities that work together to provide high quality palliative care services.[9]

D. Factors associated with the development of the payment model

In their documents and interviews, awardee leaders observed that while partnerships were helpful, challenges were the main factors associated with payment model development. As discussed, Four Seasons was unable to implement the payment model in the third program year as originally planned because of delayed claims data. One reason for the delay was that the awardee found it difficult to meet rested encryption standards for security safeguards. Awardee leaders reported one important lesson that they learned from this experience: it always takes longer and costs more than expected to acquire data. Awardee leaders also mentioned that the lack of physical functional scores in the claims data made it difficult to define appropriate patients for whom palliative care would be appropriate.

The second challenge was related to the lack of a reimbursement stream under Medicare Part B to cover palliative care services provided by an interdisciplinary team. Although physicians and nurses can bill for their clinical services, other members of the team, such as social workers, are not eligible providers under Medicare Part B. An alternative payment model would overcome this barrier. Despite these challenges, awardee leaders said that working with Duke and other key stakeholders, such as payers and experts of palliative care, facilitated the development of the payment model.

According to awardee leaders, the third and fourth challenges were the scope of the analysis and the time frame for developing a community-based palliative care payment model. More specifically, during the virtual site visit in the second year of the cooperative agreement, the Duke researchers mentioned that (1) a demonstration involving more than just Four Seasons and its implementing partners would be required to effectively test the payment model and (2) three years is too short a time to calculate the long-term impacts of a program involving end-of-life care.

[9] See https://aspe.hhs.gov/system/files/pdf/255906/ProposalAAHPM.pdf.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Four Seasons and most of its implementing sites reported that they plan to sustain the program. Some sites are also expanding the program after the cooperative agreement to new populations or by providing new services. The implementing sites said that they found that the program was valuable and aligned with their organizations' priorities, making the decision to sustain and expand the program relatively easy. However, low revenue for palliative care services challenges sustainability at all sites and may prevent one site from sustaining the program at all. Four Seasons also began finalizing agreements with an organization that is interested in replicating the program in North Carolina.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Four Seasons had committed to sustaining the program after the cooperative agreement, although the other implementing sites had not begun to plan how to sustain the program. The awardee was prepared to continue funding the program at its own site as it did for the 12 years before the cooperative agreement by using revenue from hospice and other billable services. The other four implementing sites had begun to implement the program

and were focused on implementation rather than sustainability planning as of the second program year.

C. Implementing the SSR plan: progress and changes

Sustainability. As of the third program year, Four Seasons and all but one of its implementing sites reported that they planned to sustain the program after the cooperative agreement. In addition to hospice revenue, awardee leaders said that new Medicare billing codes will help to financially support palliative care in all implementing sites. The codes are for complex chronic care management, advanced care planning, and remotely-delivered service from a clinician. Despite this favorable development, one site reported during our third round of site visits that it was unsure of how the program would be financially viable without the funding from the cooperative agreement and was therefore uncertain of its ability to sustain the program at its site after the cooperative agreement.

> "This grant was something we were going to do anyway, so if fit well with us. It just gave us a way to measure what we were not measuring before."
>
> —Leader at implementing site

The other implementing sites reported that they plan to sustain the program as is but with some modifications. Leaders from all sites mentioned that the program would probably operate with fewer clinical, administrative, and leadership staff. Interview respondents believed that having fewer staff would force them to find ways to operate the program more efficiently, but they did not express concern about not being able to perform required program activities. Moreover, all interviewees reported that Four Seasons will continue to support sites through phone calls and other technical assistance.

Scalability. Four Seasons and all but one of its implementing sites also reported that they plan to expand the program after the cooperative agreement. One site reported that it expects to increase outpatient palliative care by offering it in people's homes. Four Seasons mentioned that it has made progress on multiple initiatives that branch off of the CPC program. First, it has submitted a grant application that will fund its effort to train and educate rural providers in palliative care beyond its catchment area. Second, it joined a National Committee for Quality Assurance project that expands and evaluates the role of patient and family advisors in palliative care services. Third, it submitted a grant application to the National Institutes of Health to expand tele-health program services.

Replicability. Four Seasons used resources from the cooperative agreement to set the stage for replication. By the end of the third program year, the awardee had started finalizing agreements with a new implementing site, also located in Catawba County. The new site will implement the program after the cooperative agreement ends. Four Seasons leaders also reported that they have been communicating with other organizations that have already expressed an interest in replicating the program.

D. Factors associated with progress toward implementing the SSR plan

Four Seasons and the implementing sites made acceptable progress towards implementing the SSR plan. Several factors were associated with this progress. For example, the awardee and its implementing sites found that the program was valuable and that it aligned with their

organizations' priorities, making the choice to sustain the program fairly easy. All implementing sites reported that before participating in the cooperative agreement, their organizations either had plans to provide or palliative care services, or they were already providing similar palliative care services. From helping to standardize their workflows to collecting data in QDACT, leaders at implementing sites said that the support they received from Four Seasons boosted their capacity to provide high quality care for patients with life-limiting illnesses.

Throughout the cooperative agreement, Four Seasons understood that cost would be a large barrier to sustainability. As discussed, interviewees explained that palliative care programs are not financially viable on their own, but they can be subsidized by hospice revenues, especially because palliative care patients are more likely to transition to hospice. Awardee leaders reported that although new Medicare billing codes "do trim your losses significantly," they do not cover overhead costs.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Johns Hopkins University

April 18, 2018

Ellie Coombs (Mission Analytics Group)
Smita Patil (Mission Analytics Group)
Peggy O'Brien-Strain (Mission Analytics Group)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Johns Hopkins University received HCIA R2 funding to support the Maximizing Independence at Home (MIND) program. The purpose of the MIND program is to identify and address the unmet needs of individuals diagnosed with Alzheimer's disease or another dementia-

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Johns Hopkins University received a three-month, no-cost extension.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

related neurodegenerative disease and the unmet needs of their caregivers. Due to challenges such as fragmentation in the health care delivery system, lack of care coordination, and social isolation, individuals with dementia and their caregivers often have care needs that are unevaluated and unmet. To address this problem, the awardee partnered with two home health agencies—Johns Hopkins Home Care Group and Jewish Community Services—to coordinate care to older adults living in and around Baltimore, Maryland, who had been diagnosed with or suspected of Alzheimer's disease or another dementia-related neurodegenerative disease. The MIND program's primary goal was to delay or prevent participants from moving out of their homes and into supported living facilities or nursing homes. Additional goals included fewer ED visits and hospital admissions as a result of safer home environments and improved well-being. The individualized, in-home care management services offered through the MIND program are what make it innovative.

Johns Hopkins University received participant referrals to the MIND program through partnerships with Maryland Medicaid, managed care organizations (MCOs), medical and social service providers, and other community-based organizations that support older adults. Once individuals were enrolled in the MIND program, memory care coordinators (MCCs) worked with an interdisciplinary team—composed of a geriatric psychiatrist, an occupational therapist, and registered nurses (RNs)—to address participant needs through regular interactions and home visits. The MCCs (1) connected participants to meaningful activities, (2) educated caregivers in how to mitigate participants' risky behaviors, (3) referred participants to services to improve the safety of the home environment, and (4) worked closely with the clinical team to coordinate care (Table I.1).

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Identify and address the unmet needs of adults who have been diagnosed with Alzheimer's disease or another dementia-related neurodegenerative disease and their caregivers in order to improve health outcomes, reduce health care costs, and prevent or delay institutionalization.
Major innovation	Individualized care management and caregiver support delivered in the participant's home by MCCs with the support of an interdisciplinary team
Program components	• Care management (primary) • Patient and family engagement (secondary) • Health information technology (secondary)
Target population	Medicare beneficiaries in and around Baltimore, Maryland, who had been diagnosed with or suspected of Alzheimer's disease or a related form of neurodegenerative dementia
Theory of change/ theory of action	By addressing participants' unmet social and medical needs, supporting caregivers, and improving home safety, the program aimed to help participants stay in their homes longer.
Payment model	Separate capitated payments for MCC and clinical staff; shared savings between Medicaid and home health care agencies
Award amount	$6,387,736
Effective launch date[a]	March 2, 2015
Program setting	Participants' homes

Table I.1 *(continued)*

Program characteristic	Description
Market area	Urban and suburban
Market location	Baltimore, MD, and the surrounding area
Target outcomes	Cost savings and improved patient quality of life resulting from participants remaining in their homes longer and in a safer home environment

[a]After the initial planning period, the awardee's program became operational as of this date.

MCC = memory care coordinator.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was partly successful in implementing its program by the end of the initial three-year cooperative agreement. Although Johns Hopkins University struggled with enrollment, the awardee effectively implemented the program for those participants who were enrolled. The awardee's final enrollment was only 342 participants, which was well short of its original goal of 600 participants. The majority of program enrollment occurred in the second program year, which forced the awardee to shorten the intervention period. Despite low enrollment, John Hopkins University built a qualified and stable team, delivered needed services to participants, kept the majority of program participants engaged, and decreased the unmet needs among program participants. Finally, participating staff reported that the program had a positive effect on the delivery of care.

Impact evaluation. Due to too few treatment beneficiaries, we do not anticipate being able to conduct a rigorous impact analysis for Johns Hopkins University. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. John Hopkins University has proposed a condition-specific, population-based payment model for dementia care. The awardee is in the early stages of discussing its model with the Maryland Medicaid program. It also has conducted extensive outreach to Medicare and Medicaid MCOs and provider organizations.

Sustainability plans. Johns Hopkins University made progress on its plan to scale the program through a licensing model by contracting with external partners to conduct market assessments. The market assessments will help the awardee develop a viable licensing model for its program. Johns Hopkins University reported challenges with sustaining the program due to insufficient time to develop and implement alternative payment or licensing models. Specifically, the awardee reported challenges with retaining program staff, who took other jobs during the cooperative agreement due to uncertainty about the MIND program's future funding.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a non-clinician staff survey on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey, which was fielded from July 2016 to October 2016, achieved a response rate of 100 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items with fewer than 11 respondents, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partly successful in implementing its program. We based this conclusion on the preponderance of evidence surrounding the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

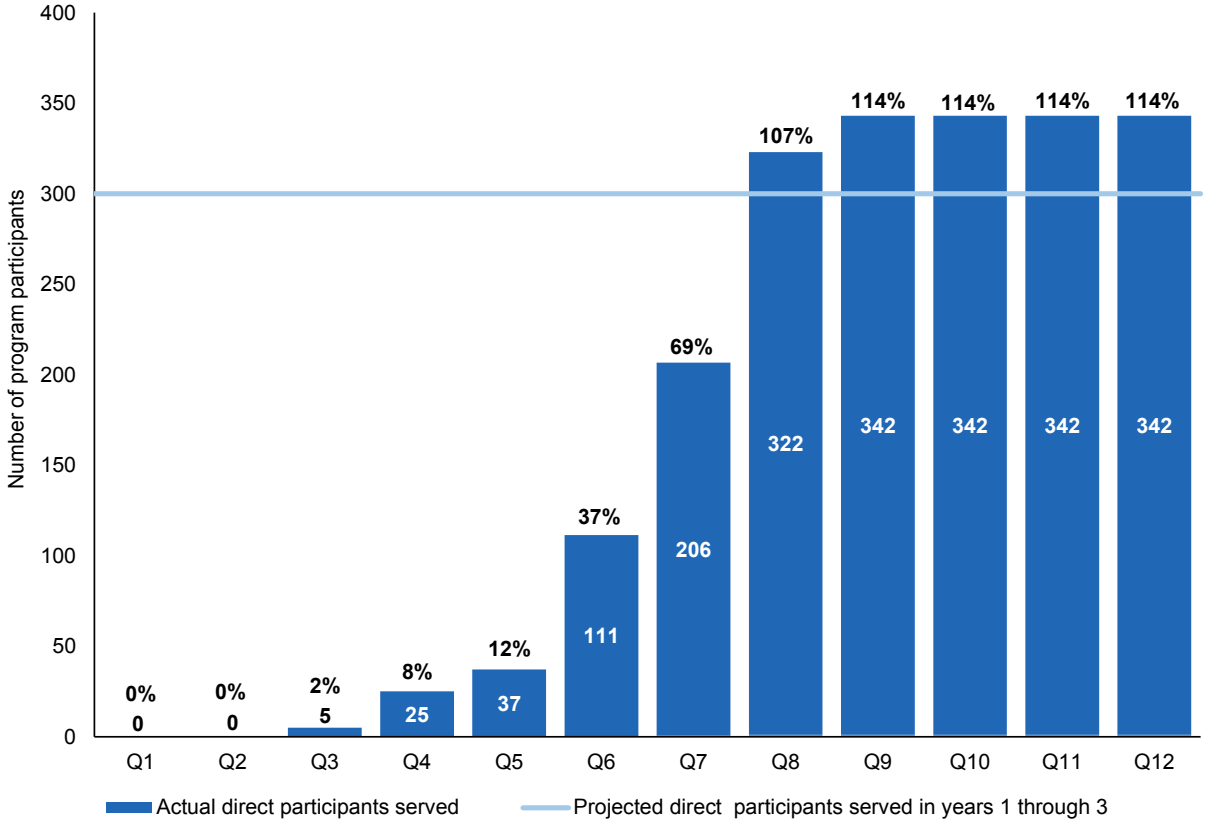
Participants were eligible for the MIND program if they lived in the Baltimore, Maryland, area; had a diagnosis of dementia (either pre-existing or diagnosed through the program); were insured through Medicare; and had a study partner (predominantly caregivers). Johns Hopkins University relaxed the MIND program eligibility criteria in the second program year by expanding the catchment area, not requiring a pre-existing diagnosis of dementia, and allowing Medicare-only participants to enroll (initially, only dual eligibles were eligible). Johns Hopkins University identified its target population through a partnership with the Maryland Medicaid program and outreach to primary care providers, Medicaid MCOs, and organizations that work with older adults. The enrollment process included an initial phone screen and an on-site visit to confirm eligibility and develop a care plan to address unmet physical health, mental health, and

social needs. This process was streamlined during the second program year by consolidating two on-site visits into one.

b. Evidence of enrollment effectiveness

Johns Hopkins University reported that it enrolled 342 participants from March 2015 (when it launched its program) through September 2016 (when program enrollment ended). Although total enrollment exceeded the three-year target (Figure II.1), the awardee had revised its target downward after the first program year, from 600 participants to 300 participants, due to slow enrollment. When measured against the three-year projection set at the end of Year 1 (enough time for the awardee to gain a better understanding of the number of eligible participants but before the awardee began planning for the end of their cooperative agreement), the awardee met 57 percent of its projection.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee lowered its enrollment target from 600 to 300 participants at the start of Year 2.

In the first program year, Johns Hopkins University experienced a setback to enrollment because a large share of recruited individuals did not meet the original MIND program eligibility

criteria. According to the awardee's self-monitoring data, 733 individuals were screened for program enrollment in the first program year, but only 185 of those individuals, or 25.2 percent, were eligible. The awardee reported that the most common factor for ineligibility was that the individuals were not dual eligibles.

The majority of enrollment occurred during the second half of the second program year, after Johns Hopkins University changed the eligibility criteria and ramped up its enrollment efforts. According to self-reported monitoring data, 230 of the awardee's 342 program participants, or 67.3 percent, were enrolled between March 2016 and September 2016. Although the share of recruited individuals that met the program eligibility program was still low, it was higher than in Year 1 (39 percent in Year 2 versus 25.2 percent in Year 1). Slow enrollment in the first program year prompted Johns Hopkins University to shorten the intervention period from 26 months to 18 months. However, the shorter intervention period was still sufficient for participants to receive the desired exposure to the intervention.

c. Barriers and facilitators associated with enrollment effectiveness

Johns Hopkins University's progress in meeting its three-year enrollment goal was influenced by several factors. First, the awardee felt that the MIND program's original target population was difficult to recruit because of the severe social isolation that is common among dual eligibles with dementia. This prompted the awardee to include Medicare-only beneficiaries in the program, vastly expanding the pool of eligible recruits. The awardee also expanded its catchment area to include additional counties surrounding the Baltimore metropolitan area. The results of expanding the catchment area were limited, however, because many of the individuals who were referred spoke Spanish only and therefore were not eligible for the program.[4]

In addition, Johns Hopkins University was hampered by the low dementia detection rates among health care providers. The awardee originally required program participants to have a prior dementia diagnosis. However, Johns Hopkins University discovered that dementia tended to be underdiagnosed among the patient population living in Baltimore and the surrounding counties. Thus, MIND program leaders decided to allow individuals with suspected dementia to enroll in the program in an effort to boost program enrollment. Individuals who displayed symptoms of dementia were able to call the MIND program referral line and go through the initial phone screen. These individuals were then assessed by a MIND geriatric psychiatrist and received a formal diagnosis when applicable.

The awardee also struggled with the problem of oversaturation of research opportunities and a wariness of research and the consent process within the community. Because Johns Hopkins University is a large research institution that conducts numerous studies and randomized trials, individuals who live in the surrounding community are frequently offered opportunities to participate in research and, as a result, feel indifferent toward becoming involved. Further, controversial research conducted by Johns Hopkins University in the past resulted in some community mistrust of the research and consent process.

[4] Johns Hopkins University did not have the resources to hire interpreters or conduct the institutional review board approval process for translated materials.

In addition to expanding MIND program eligibility in the second program year to improve program enrollment, Johns Hopkins University also increased the number of partnerships it used for participant referrals. The awardee executed a massive, three-stage mailing campaign to potentially eligible individuals through the Maryland Medicaid program, which led to a significant uptick in program enrollment. Johns Hopkins University also executed a smaller mailing campaign through an MCO. Further, the awardee established partnerships with a greater number of medical and service providers in the program catchment area as well as with community-based organizations, such as churches and elder adult community centers.

2. Delivery of program services

a. Description of and changes to the service delivery model

The MCCs helped participants reduce unmet social, medical, and safety needs by connecting them to meaningful activities and services; educating caregivers in how to mitigate participants' risky behaviors; and coordinating social and medical services by reminding participants to schedule necessary appointments, arranging transportation, and monitoring adherence to medication regimens. By keeping participants safer and happier in their homes, the MIND program attempted to delay or prevent participants from moving into institutions and to reduce ED visits and hospitalizations.

> One MCC described an interaction with a participant who liked to take the bus around the city "for fun" but often got lost. The MCC worked with him to obtain a Medic Alert ID bracelet. In addition, she helped him gain access to Meals on Wheels five days a week and a paid caregiver two days a week.

In the last program year, Johns Hopkins University expanded the array of services offered through the MIND program by adding tele-health consultations, occupational therapy, and a monthly newsletter. Tele-health connected participants and their caregivers with a geriatric psychiatrist to discuss clinical needs and next steps in care. Johns Hopkins University also piloted occupational therapy services through its partnership with the Tailored Activity Program (TAP), a home-based occupational therapy intervention for individuals with dementia and their caregivers. Although the services provided through TAP were beneficial to some MIND participants, the awardee felt that the services were not a good fit for the more intensive needs of MIND participants. Finally, the monthly newsletter helped area organizations that work with older adults to share information about related resources and events.

b. Evidence of service delivery effectiveness

Johns Hopkins University was effective in implementing the MIND program, as demonstrated by the frequency and mode of services provided, high rates of participant retention, and, most importantly, a decrease in participants' unmet needs. Johns Hopkins University was also successful in engaging and training staff and partner organizations. Below, we describe in more detail evidence for the effectiveness of the delivery of MIND program services.

Delivery of intervention services. The MCCs worked with participants and caregivers through regular home visits, telephone calls, text messaging, and tele-health. In a given month, the MCCs made contact with participants four or five times. As shown in Table II.2, telephone calls were the most frequent method of contact, with approximately three calls per enrollee per month during the last program quarter. Emails were also somewhat frequent, with almost one contact per month. The MCCs also visited participants in their homes and sent text messages

approximately once a quarter. In the last program quarter, John Hopkins University conducted 15 tele-health consultations.

Table II.2. Frequency of participant contacts in the MIND program

	Frequency		
Mode of Communication	August 2015 (end of Q4)	August 2016 (end of Q8)	August 2017 (end of Q12)
Telephone	4.4	3.2	3.3
In person	3.8	0.9	0.3
Email, mail, or fax	0.7	1.1	0.8
Text	0.0	0.1	0.3

Source: Johns Hopkins University Self-Monitoring Measurement Plan, 12th quarter.

Although Johns Hopkins University did not report the distribution of contacts across participants, interviewees reported that some participants needed more attention than others. However, they said that participants' needs also changed over time as they faced and resolved various challenges or crises. The MCCs reported that they met participants where they were at, both in terms of the mode and frequency of contacts. Some caregivers worked full-time, so the MCCs had to be flexible with the timing of their communications and use less invasive modes, such as text messaging. Other participants were going through a medical crisis and needed the MCCs' frequent support in communicating with clinicians or addressing hospital discharge needs.

A decrease in the number of unmet needs over time indicates that these MCC contacts paid off. John Hopkins University conducted needs assessments at the initial enrollment visit, the 9-month visit, and the 18-month visit. Table II.3 presents the average number of unmet needs per enrollee at baseline and the percentage decrease at 18 months. At baseline, enrollees experienced 3.5 personal safety needs on average, ranging from wandering to having a throw rug that was a fall risk. Through interventions, such as providing advice, resources for home modifications, and tracking bracelets, these unmet needs decreased by 44 percent. Unmet needs related to meaningful activities also decreased significantly—by 58 percent. The MCCs supported participants by enrolling them in Medicaid-funded day programs or connecting them to the occupational therapist. One MCC reported that a participant's caregiver felt that her mother was too low functioning to engage in any activities. However, with gentle nudging from the MCC and occupational therapist, the participant greatly benefited from word searches and yarn crafts.

Table II.3. Percent decrease in unmet needs, from baseline to 18 months

Unmet need	Average number of unmet needs per enrollee, baseline	Average number of unmet needs per enrollee, Q12	Percent decrease from baseline
Cognitive symptoms	0.5	0.10	-19%
Neuropsychiatric symptoms	1.5	0.53	-35%
Personal safety	3.5	1.54	-44%
Medical care	2.3	0.64	-28%
Meaningful activities	2.1	1.22	-58%
Legal issues/advanced care planning	1.7	0.51	-30%
Care financing	0.4	0.01	-24%

Source: Johns Hopkins University Self-Monitoring Measurement Plan, 12th quarter

Staffing and training. The MIND project team included Johns Hopkins University leaders (project director and project manager) and clinical staff (geriatric psychiatrist, RNs, and an occupational therapist), in addition to the MCCs contracted by the partner home health agencies. Although the Johns Hopkins University team remained stable throughout the cooperative agreement, the number of MCCs required to staff the program increased as more participants enrolled in the program. At the MIND program's peak enrollment at the end of the award's second year, there were approximately 23 full-time staff, including MCCs, research assistants, and the clinical team. The program's quarterly retention rate remained at 90 percent or above throughout the award. The MCCs who left the program primarily did so because of uncertainty surrounding the no-cost extension and their job security.

The MCCs received considerable training through the MIND program, including participating in a formal course on the assessment process; available community resources; and use of the Dementia Care Management System (DCMS), which tracked client interactions and care needs. New MCCs also shadowed more experienced MCCs to gain tips on client interactions and identify available services and supports that could address client needs. Finally, the MCCs and the Johns Hopkins University team participated in weekly meetings to share experiences and troubleshoot participant issues. Although these meetings primarily served as a mechanism to advance a participant's care plan, they also helped the MCCs learn more about caring for people with Alzheimer's and dementia. Survey responses collected prior to the last round of interviews were consistent with interview respondents' observations on training. Of the 19 staff members who completed a survey on their experience with the MIND program, 90 percent participated in some type of formal training. In addition, over 80 percent reported that they received informal instruction through multiple venues, including staff meetings, mentoring, shadowing, and individual and group supervision. The vast majority of staff (80 percent) reported that the training helped them learn new skills and improve job performance.

Engagement of providers and provider organizations. By partnering with two home health agencies, Johns Hopkins University simulated a real-world environment. Most likely, a Mind model would operate out of one of these agencies, instead of a large hospital system such as Johns Hopkins University's hospital system. However, these agencies did not play a large role in the development and implementation of the MIND program. Instead, they were responsible for contracting the MCCs, while the majority of the training and oversight fell to Johns Hopkins

University. That said, home health agency leaders reported that the Mind program was a natural fit for their organizations. They participated in regular meetings with Johns Hopkins University to stay abreast of programmatic changes.

Engagement of program participants. Few participants withdrew prematurely from the program, which Johns Hopkins University interpreted as an indicator of successful program engagement given the considerable length of the intervention period (18 months). Over the three-year cooperative agreement, 137 participants were disenrolled, primarily because they completed 18 months of the intervention (29 percent); transitioned permanently into a higher level of care (for example, a nursing facility) (27 percent); or died (31 percent). In total, 12 percent of participants withdrew from the intervention. None of the participants were lost to follow-up.

Participant engagement also refers to how actively participants communicated with their MCCs and utilized program services. As previously mentioned, Johns Hopkins University did not report distributions on the frequency and scope of contacts. However, in interviews, multiple MCCs and Johns Hopkins University staff members reported that about 20 percent of participants were not fully engaged in the program. About half of these were so high functioning that they were able to manage independently and did not require program services. The other half was on the opposite end of the spectrum. As one MCC reported, the care needs of those participants were so intensive that they simply did not have time to properly engage with the program, despite the fact that high-needs participants stand to benefit the most from receiving the MIND program intervention.

c. Barriers and facilitators associated with service delivery effectiveness

Johns Hopkins University faced few barriers to program implementation. Instead, the program benefited from qualified and engaged staff, an innovative spirit, and an effective case management system. Perhaps the greatest asset to the MIND program was the MCCs themselves. The MCCs were valued for being resourceful, persistent, and not becoming overwhelmed by the challenges their participants faced, but instead prioritizing and calmly tackling issues one at a time.

> "I can think of a client who kind of went into hiding for a while. He would not return phone calls. I would mail him things occasionally. But I kept at it and when his study partner finally got back to me, I learned he had been in a very serious depression and was not able to return phone calls, and he was so glad that I persisted."
>
> —MCC

The MCCs greatly benefited from the expertise and frequent communication with the interdisciplinary team, including the geriatric psychiatrist, the RNs, and the occupational therapist. During weekly, two-hour clinical team meetings, the MCCs went over changes in the participants' dementia management, recent activities to address unmet needs, challenges that arose during the previous week, and potential strategies to address them.

Quality improvement and innovation were also foundations of the program. Johns Hopkins University leaders and staff continuously evaluated program success and developed innovative strategies to better serve participants. Most importantly, they developed new interventions to reduce high rates of ED use and hospitalizations among participants. Through tracking ED visits and hospitalizations, the MCCs could visit participants in the hospital in a timelier manner and ensure that they had the supports they needed when they returned home. Moreover, the awardee

increased communication between MIND clinical staff and caregivers of participants who had been hospitalized (or were at risk of hospitalization) to discuss risks, such as falls, dehydration, and medication mismanagement. Tele-health consultations, implemented in the third year of the program, also allowed clinicians to assess firsthand a participant's needs and to provide advice to caregivers.

Finally, the DCMS was a key component of program success. Built specifically for the MIND program, this health information technology system had two main features. First, it contained a repository of informational and service-related resources for the MCCs to access. As the MCCs served more participants and identified more resources, they populated the database with organizations, service providers, and tips for treating dementia symptoms and behaviors. This process allowed the MCCs to systematically share resources with each other, which improved their effectiveness. Second, the DCMS stored participants' information, which allowed the MCCs to track their needs, progress, and outcomes. As the MCCs worked with the participants, they indicated in the DCMS whether a need was met or partially met.

C. Assessment of perceived program effects on the delivery of care and outcomes

Of the 22 Johns Hopkins University staff members and MCCs who completed the survey, the vast majority reported that the MIND program was worth the effort (90 percent) and was either very effective or somewhat effective in achieving its goals (95 percent). All but one respondent agreed that the program was making a difference in meeting critical needs in the community. Most commonly, staff reported that the program provided important supports that reduced ED visits and institutionalization. At this point, Johns Hopkins University has completed only a preliminary analysis on the impacts of the MIND program on institutionalizations and ED visits. However, interviewees said that the initial data appeared to indicate a decrease in health care costs among MIND participants compared to a control group.

> "No other program provides support to the caregiver the way MIND does. The acknowledgement of the work a typical caregiver does is important for their health and well-being. The likelihood of placement [in an institution] is high if they burnout."
>
> —MCC

The Johns Hopkins University staff and MCCs reported that the MIND program helped participants stay at home longer and use the ED less often due to caregiver emotional support and identification and referrals for treatment of common medical issues. Multiple interviewees emphasized the importance of supporting caregivers so that they did not rely on the ED for respite care when they were no longer emotionally capable of caring for their loved ones. A member of the John Hopkins University clinical team recounted an instance in which she provided advice through tele-health to a caregiver about reducing her husband's aggressive and agitated behaviors. The caregiver decided not to take her husband to the ED, and instead learned to better treat him at home.

Staff also felt that by connecting participants to medical care, the MIND program was generating positive effects. Johns Hopkins University staff recognized early on that participants were going to the ED or being hospitalized due to avoidable illnesses, such as dehydration or urinary tract infections. Therefore, John Hopkins University trained the MCCs on signs of these

common conditions, so they could refer patients to their primary care physicians or obtain advice from the Johns Hopkins University clinical team. The tracking of hospitalizations for discharge planning and follow-up were also perceived as having positive effects on participants' health and thus reductions in readmissions.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Johns Hopkins University staff and the MCCs were confident that the program positively impacted participants and their caregivers. Through education and connections to community resources, the MIND program reduced participants' unmet needs related to neuropsychiatric symptoms, personal safety, medical care, meaningful activities, legal issues, and finances. They also reported that caregivers experienced less burnout due to the emotional support provided by the MCCs.

Although staff believed that the program had the potential to reduce the cost of care to CMS due to delayed institutionalizations and reduced ED visits and hospitalizations, they were concerned that impacts might be limited because of the shortened intervention period and inclusion of Medicare-only beneficiaries. Medicare-only beneficiaries tend not to be as high need as dual beneficiaries. In addition, because Medicaid, not Medicare, covers the costs of long-term care, delayed institutionalization for Medicare-only beneficiaries will not generate savings for Medicare. Further, because of slow enrollment during the first program year, the awardee extended the enrollment period and shortened the intervention period from 26 months to 18 months, thus potentially reducing potential program impacts.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Johns Hopkins University's MIND program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data. Johns Hopkins University received a three-month no-cost extension, and its program ended November 30, 2017. Enrollment into the program ended September 2016.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: Johns Hopkins University

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 14-24 months of program exposure at program end	249[a]
Projected Medicaid population with 6 months of program exposure at program end	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	2,402
Likelihood of all-cause hospitalizations	1,361
MDE sample size requirement to detect 20% effect	
Total expenditures	601
Likelihood of all-cause hospitalizations	340
Participation/Selection bias of concern	Yes, patient self-selection high or high refusal rate
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Questionable, moderate rate of identification of treatment group with claims with some dilution of treatment effect within an intent-to-treat framework
Likelihood of solid comparison group	Some issues, but probably surmountable
Do claims identify the primary expected effects?	Some effects observed in claims data, but important effects likely missing
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Too few treatment beneficiaries for primary outcomes
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys

Table III.1 *(continued)*

Evaluability domain	Response
Implementation data that will be analyzed	Awardee will have extensive 9-, 18-, and 24-month assessments on participants who remain at home (with 18 months as the full dose for program exposure). Assessments include cognitive function, functional dependency, depression, patient quality of life (at 18 months only), neuropsychiatric behavior, caregiver burden, and care satisfaction (18 months only).

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We do not expect to have sufficient statistical power to evaluate the MIND program because of low enrollment. Overall enrollment is too low to have sufficient statistical power for core measures. The likelihood of all-cause hospitalizations is closest to meeting the threshold for sufficient statistical power. However, the program is not expected to have a major impact on this measure. The low share of Medicare-Medicaid dual eligibles (70 percent of participants) further reduces the statistical power for assessing the primary effect of delaying the time to nursing home placement and its associated costs. By the end of its no-cost extension in November 2017, the awardee projected that 50 percent of participants would have received the full program exposure of 18 months. Absent a rigorous impact analysis, we will explore with the awardee the availability of data that it is collecting at three points during its model testing period (baseline, 9 months, and 18 months) from participants and their caregivers who remain at home. If the data are available and complete, we will report on changes over time in cognitive function, functional dependency, depression, patient quality of life (at 18 months only), neuropsychiatric behavior, caregiver burden, and satisfaction with care (18 months only). Participants who remain at home may also receive a 24-month assessment, but only 10 percent are projected to reach the 24-month point by November 2017. We will also report on the experiences of staff and participants, based on our surveys, and assess the feasibility of studying the delay to nursing home among program participants relative to a comparison group.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents our summary of the characteristics of the treatment group, which we measured for the year before the intervention began for each beneficiary. Recruitment for the treatment group was conducted through letters to Medicaid beneficiaries, advertising, and referrals from community partnerships or community meetings. Approximately 10 percent of those identified or referred to the program declined to participate. There was a larger drop-off through the screening process. Johns Hopkins University originally targeted only individuals enrolled in Medicare and Medicaid (that is, dual eligibles) for the MIND program. However, approximately 60 percent of those screened were excluded because they were not dually eligible. Another 20 percent were ineligible for other reasons, including not having dementia or not living in the catchment area. Overall, about one in five individuals screened were eligible and about two-thirds of those who were eligible were enrolled. To increase enrollment, Johns Hopkins University expanded the MIND program's eligibility criteria to include Medicare-only beneficiaries, starting in early 2016. The treatment group in this analysis partially reflects the shift in eligibility criteria.

Johns Hopkins University began enrolling beneficiaries in the MIND program in March 2015. When enrollment ended on September 30, 2016, the program had 348 participants.[5] Seven of these enrollees later withdrew consent.

For the purpose of presenting baseline characteristics in this report, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare fee-for-service (FFS), both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who had met all program criteria for a period of 90 days during the baseline year (the 365 days immediately prior to their enrollment). The calendar period covered by the baseline quarters is based on the enrollment date for each participant and therefore varies by participant. After we excluded beneficiaries who did not meet the above criteria, 249 participants were included in the analysis of baseline characteristics for this report. This is the final analytic sample for this evaluation.

Table III.2 shows the demographic characteristics of MIND participants. Nearly all of the participants (94 percent) are 65 or older; 37 percent are 85 or older. The participants are predominantly female (77 percent) and black (60 percent). Most participants (78 percent) became eligible for Medicare through age or survivor benefits. Of the 249 beneficiaries included in the analysis, 175 beneficiaries (70 percent) are dual eligibles. MIND participants have substantial health care needs. The average hierarchical condition categories (HCC) risk score for MIND Medicare beneficiaries is about twice the average Medicare FFS HCC score—indicating that they are expected to have much higher costs than the average Medicare FFS beneficiary.

[5] The enrollment data provided by the implementation and monitoring contractor and awardee self-reported enrollment data cite a final enrollment figure of 342 participants, however, the finder file enrollment data cites a final enrollment figure of 348 participants, with 7 participants having later withdrawn consent. The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Johns Hopkins University's program through September 30, 2016

	All participants (N = 249)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	15	6
65 to 74	47	19
75 to 84	92	37
85 and older	95	38
Gender		
Female	185	74
Male	64	26
Race		
White	77	31
Black	149	60
American Indian, Alaska Native, Asian/Pacific Island American, or other	17	7
Hispanic	4	2
Original reason for Medicare eligibility		
Old age and survivor's insurance	194	78
Disability insurance benefits	54	22
End-stage renal disease (ESRD)[a]	1	0.4
Hospice[b]		
Medicare/Medicaid dual status, percentage dual[b]	175	70
HCC score[c]		Statistic
Mean		1.99
25th percentile		1.05
Median		1.58
75th percentile		2.53

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of March 31, 2017.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the beneficiary consented to be enrolled in the program. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

The substantial health care needs of MIND participants are reflected in their high level of baseline health care expenditures and utilization. Table III.3 lists a common set of utilization and cost measures, including core measures from CMMI. By addressing unmet care needs in the home and reducing caregiver burden, Johns Hopkins University attempted to reduce transitions to long-term nursing facilities—which is a cost to Medicaid for dual eligibles. These costs are not included in Table III.3. However, the awardee also expects the MIND program to shorten inpatient stays and, in turn, lower health care utilization expenditures among participants. We examined baseline cost of care by calculating average per beneficiary per month (PBPM)[6] Medicare expenditures, in total and by major types of services. The total average PBPM Medicare expenditure during the baseline year was $1,892—more than twice the 2014 national average for Medicare FFS beneficiaries of $792.[7] The average PBPM Medicare expenditure in the baseline year varied substantially from quarter to quarter, ranging from a low of $1,363 in quarter 1 to a high of $2,231 in quarter 2. Medicare expenditures for inpatient services ($729 PBPM) were the largest driver of total cost of care for participants. However, because of the Medicare billing practices in Maryland, the inpatient expenditures likely include psychiatric services and other non-acute inpatient stays, resulting in an inflated figure for acute care inpatient service expenditures. The next highest expenditures for MIND participants were for outpatient services ($379 PBPM) and physician services ($321 PBPM).

For participants, the annual hospitalization rate of 688 per 1,000 beneficiaries was well above the national annual rate of 274 admissions per 1,000 beneficiaries in 2014, though this figure is also likely inflated for the reason stated above. Readmission rates were not particularly high, at 14 percent of all discharges. The annual rate of 1,011 ED visits that did not lead to hospitalization per 1,000 participants was more than double the 2014 national Medicare FFS rate of 454 per 1,000 beneficiaries. Furthermore, the annual observation stay rate of 220 per 1,000 participants was nearly four times the observation stay rate of all Medicare FFS beneficiaries in 2013 (60 per 1,000 beneficiaries).[8] Approximately 35 percent of MIND participants had at least one hospitalization during the year prior to enrollment, 40 percent had an ambulatory ED visit at least once, and 21 percent had at least one observation stay. Outside of inpatient and outpatient utilization, MIND participants were more likely to have specialist visits compared to primary care visits—10,135 specialty services used per year per 1,000 participants compared to 7,593 primary care visits per year per 1,000 beneficiaries.

Over the course of the four baseline quarters leading to enrollment, we observed no discernable trend in the average PBPM total payment nor in the average PBPM payments for inpatient, outpatient, and physician services. Overall, the first quarter showed substantially lower levels of inpatient, outpatient, and skilled nursing services compared to the other quarters, while the second quarter showed higher utilization for most services.

[6] The months referred to in our calculations are 30-day periods rather than calendar months.

[7] Except for ambulatory observation stays, the national cost and utilization data presented here and in the next paragraph are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

[8] MedPAC. "A Data Book: Health Care Spending and the Medicare Program." Washington, DC: MedPAC, June 2015.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Johns Hopkins University's program through September 30, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	249	241	246	247	249
Average Medicare expenditures PBPM[a]					
Total	1,892 (213)	1,363 (195)	2,231 (477)	2,146 (362)	1,685 (216)
Acute inpatient	729 (122)	511 (106)	927 (239)	816 (253)	575 (133)
Inpatient other[b]	6 (6)	0 (0)	24 (24)	0 (0)	0 (0)
Outpatient[c]	379 (56)	221 (39)	491 (182)	451 (128)	348 (54)
Physician services	321 (32)	271 (30)	338 (45)	345 (64)	316 (36)
Home health	181 (19)	131 (28)	182 (33)	143 (27)	258 (38)
Skilled nursing facility	224 (42)	179 (67)	197 (63)	345 (88)	151 (57)
Hospice	14 (12)	26 (21)	20 (20)	10 (10)	0 (0)
Durable medical equipment	37 (7)	24 (5)	51 (32)	36 (9)	38 (8)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	688 (82)	571 (102)	967 (240)	601 (105)	580 (103)
Outpatient ED visits[e]	1,011 (98)	773 (124)	1,082 (156)	1,071 (171)	1,111 (175)
Primary care visits in any setting	7,535 (547)	6,318 (720)	7,543 (663)	8,424 (989)	7,745 (745)
Primary care visits in ambulatory settings	5,594 (421)	4,604 (453)	5,133 (438)	6,460 (820)	6,086 (596)
Specialist visits in any setting	11,437 (737)	10,182 (756)	13,217 (1,426)	11,979 (1,177)	10,176 (761)
Specialist visits in ambulatory settings	8,391 (487)	7,897 (613)	8,560 (684)	8,668 (746)	8,373 (592)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	37 (3)	13 (2)	18 (2)	13 (2)	12 (2)
Percentage with an outpatient ED visit[e]	50 (3)	16 (2)	21 (3)	19 (2)	20 (3)
Percentage with a 30-day readmission among all discharges	14 (3)	7 (5)	13 (5)	18 (6)	13 (6)
Percentage of participants with a readmission among all participants	7 (2)	1 (1)	2 (1)	3 (1)	2 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of March 31, 2017.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

John Hopkins University has proposed a condition-specific, population-based payment model for dementia care. The awardee is in the early stages of discussing its model with the Maryland Medicaid program. It also has conducted extensive outreach to Medicare and Medicaid MCOs and provider organizations.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Under the awardee's proposed payment model, the fee for MIND services would be $250 per month for each program participant, split between the MCCs (located within a home health agency) and the clinical team (located within Johns Hopkins University), until the participant enters a nursing facility or dies. These payments would cover the costs of care management activities, including the MCCs' contacts with participants via phone, in person, or through tele-health, as well as clinical services provided by the clinical team of geriatric psychiatrists, RNs, and occupational therapists. The team is not at financial risk for other services a participant may receive.

Johns Hopkins University based the payment amount on its assessment of the staff costs of providing services to an average caseload of approximately 42 patients, though its leaders expected these costs could vary based on changes in caseload or the intensity of services required by patients. The awardee reported that it is conducting analyses to figure out how to risk adjust the payment based on key characteristics of Medicare-Medicaid dual eligibles with dementia. Eventually, the awardee hopes to incorporate shared savings with payers—similar to the Comprehensive Primary Care model—so that home health agencies have the opportunity to share the savings that accrue as a result of delaying participants' entry into a nursing home.

Medicare and Medicaid are both potential payers for this model, as well as risk-bearing organizations such as accountable care organizations (ACOs), Medicare Advantage plans, or Medicaid MCOs. For Medicare, Medicare ACOs, the Medicare Advantage plan, and Medicaid MCOs, the incentive in reimbursing or providing MIND services would be to reduce costs associated with hospitalizations, readmissions, and outpatient care. None of these payers would be responsible for costs associated with nursing entry. In contrast, delaying or preventing entry into the nursing home is a significant incentive for Medicaid to cover MIND services.

C. Status of the payment model

Because Medicaid is the payer most likely to benefit financially from delayed entry into nursing homes, Johns Hopkins University has engaged with the Maryland Medicaid program throughout the cooperative agreement to encourage adoption of the MIND program. At the time

of our interview, the awardee was conducting analyses based on dually enrolled beneficiaries by using Medicaid claims to determine whether the program had resulted in Medicaid cost savings due to delayed or reduced long-stay nursing home admissions and lower nursing home facility expenditures. Johns Hopkins University was also obtaining Medicare claims to determine whether the program had reduced expenditures on Medicare inpatient or ambulatory care.

Johns Hopkins University has had conversations with organizations such as Medicare Advantage plans, Medicaid MCOs, and ACOs that are at financial risk for the care of Medicare beneficiaries or Medicaid enrollees; we discuss this in Chapter V.

D. Factors associated with the development of the payment model

The benefits of the MIND program for Medicare are hypothesized to be the impact on hospital and outpatient costs. However, Johns Hopkins University reported that delays in receiving access to Medicare claims data limited the ability to conduct analyses to identify potential Medicare cost savings, which created a barrier to developing the payment model. The awardee also noted that the cooperative agreement ended before some participants received the full dose of the intervention, which likely will affect the ability to detect time-sensitive changes in spending patterns and cost savings over time. Finally, awardee leaders noted that a Medicare care management billing code specific to dementia patients would be helpful to their program because they said existing codes do not adequately reimburse for the intensity of care management services required by these patients.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Johns Hopkins University made progress on its plan to scale the program through a licensing model by contracting with external partners to conduct market assessments. The market assessments will help the awardee develop a viable licensing model for its program. The awardee reported experiencing challenges with sustaining the program due to insufficient time to develop and implement alternative payment or licensing models. Specifically, the awardee reported challenges with retaining program staff, who took other jobs during the cooperative agreement due to uncertainty about future program funding.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Johns Hopkins University was looking at funding options to sustain the program. The awardee had met with the state Medicaid program to explore reimbursement options, received Medicaid claims data to inform the payment model, and begun collaborating with home health agencies that provide program staff.

To scale the program, Johns Hopkins University was developing online educational modules and vetting potential investors who would lease the program to others. The awardee was also

considering using HCIA R2 funds to conduct a market assessment to inform a strategy to rapidly disseminate the program.

C. Implementing the SSR plan: progress and changes

Sustainability. In the third program year, Johns Hopkins University explored ways to fund the program to make it sustainable. Strategies the awardee was considering included developing the payment model described in the previous chapter, combining the MIND program with another program, or continuing the program as a clinical research trial or pilot study.

Scalability. Johns Hopkins University also continued efforts to scale the program by monetizing it through licensing agreements and creating online educational modules. First, to explore the licensing option, the awardee hired a company to conduct a market assessment, which would inform how to make a program model that could be licensed, sustainable, and scalable. Awardee leaders also hired a consultant to conduct a separate market assessment to understand which markets have the lowest barrier to adopting and sustaining the program. Second, Johns Hopkins University began beta testing a comprehensive web-based training package that would prepare interested sites for implementation of the MIND service delivery model. The web-based training package was developed during the first two program years. The training is designed for self-guided learning of existing program documents and resources—such as training materials, protocol manuals, and assessment tools—in addition to lessons learned from the cooperative agreement. According to awardee leaders, the training is customized to provider type. It includes a 40-hour, web-based certificate program for care coordinators and 16-hour training on MIND service areas for geriatric psychiatrists, RNs, and occupational therapists.

Replicability. Johns Hopkins University reported no plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

Awardee leaders believe that the market assessments will help Johns Hopkins University develop a program model that can be scaled and sustained through licensing agreements. They said that not securing funding challenged their ability to sustain the program, particularly with regard to retaining program staff through the end of the cooperative agreement. Interviewees said that the cooperative agreement was not long enough to develop the alternative payment model or a licensing model. Multiple program staff took new jobs during the cooperative agreement, knowing that their jobs were unlikely to be sustained because of the lack of future funding.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Icahn School of Medicine at Mount Sinai

April 18, 2018

Suzie Witmer (Mathematica Policy Research)
Kristin Geonnotti (Mathematica Policy Research)
Jia Pu (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Icahn School of Medicine at Mount Sinai used funding from HCIA R2 to implement a Hospital at Home (HaH) program and develop an associated payment model. A funding

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

mechanism for the HaH model does not currently exist and Mount Sinai hopes to advance the national spread of the HaH program by developing a payment model.

Mount Sinai originally proposed that the Mobile Acute Care Team (MACT) program would provide both acute and post-acute care in patients' homes. Through the three-year cooperative agreement, Mount Sinai made extensive program modifications to increase enrollment and enhance care provided to participants. By the end of the cooperative agreement, the MACT program evolved to also provide subacute rehabilitation[3] at home (RaH), palliative care, observation unit care,[4] pediatrics care (as a pilot program), and care for patients who are averse to entering the hospital. Mount Sinai considered HaH to be the umbrella under which acute care, observation unit care, palliative care, and hospital-averse care were provided.[5] Mount Sinai considered RaH to be a separate program arm. The MACT program was intended for adults (age 18 and older) who lived at home in Manhattan, were enrolled in certain insurance plans, had a safe home environment with access to a single (not shared) bathroom, and had sufficient support for or the ability to complete necessary daily activities (such as cooking meals or using the phone to call for help). The criteria varied slightly for the pediatrics pilot and are not documented here. The HaH and RaH arms used additional inclusion and exclusion criteria specific to each arm.[6] These criteria are listed below.

- Inclusion criteria
  - The HaH arm's additional inclusion criteria included that the patient must (1) be enrolled in Medicare fee-for-service (FFS), Healthfirst Medicare, Healthfirst Medicaid managed care, or the Healthfirst health maintenance organization and (2) meet the Milliman Care Guidelines (MCG) for hospital admission, a set of evidence-based guidelines that support decision making related to admitting patients to the hospital.
  - The HaH arm also had specific medical condition criteria, which indicated that a patient must have at least one of the following eight conditions: (1) congestive heart failure (CHF) or heart failure, (2) chronic obstructive pulmonary disease (COPD) or asthma, (3) dehydration, (4) diabetes, (5) pneumonia, (6) cellulitis, (7) urinary tract infection, or (8) pulmonary embolism or deep vein thrombosis (DVT). At about the midway point of the cooperative agreement, these diagnostic requirements were loosened.
  - The RaH's additional eligibility requirements included that (1) the patient must be enrolled in Medicare FFS or be dually eligible and (2) the patient's physical therapist,

[3] Medical professionals sometimes use the term "subacute rehabilitation" to refer to short-term programs conducted in a nursing home (see "Your Guide to Hip Replacement Surgery" at www.mountsinai.org/static_files/MSMC/Files/Faculty%20Profile%20Pdfs/Calin%20Moucha/Hip%20Replacement%20Brochure.pdf). Subacute rehabilitation is thus equivalent to the services provided under the Medicare Skilled Nursing Facility (SNF) benefit, which, in turn, is one of Medicare's post-acute benefits.

[4] Starting in Year 3, Mount Sinai began categorizing the observation unit subgroup as its own arm. However, because it was historically captured under the HaH arm and its implementation was similar to the HaH arm, we categorize it as part of the HaH arm in this report.

[5] This narrative speaks broadly to the two high-level groupings: HaH and RaH. If any information is specific to the HaH subgroup, the subgroup is specifically mentioned.

[6] Mount Sinai revised the HaH arm's eligibility criteria during the program to expand the eligible population.

physician, or social worker must indicate that the patient needs subacute rehabilitation services.

- Exclusion criteria
  - There were two sets of HaH exclusion criteria: (1) those that were applicable to all conditions and (2) those that were condition-specific. The all-condition exclusion criteria were generally based on needs for clinical services, such as critical care. Condition-specific exclusion criteria were very detailed. For example, for patients with heart failure, the exclusion criteria included having a recent definite or suspected myocardial infarction or pulmonary embolism.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The MACT program provides acute, post-acute, observation unit, and subacute rehabilitation services in a patient's home.
Major innovation	Hospital at Home (HaH) payment model Expansion of HaH model to provide services to new patient populations, including patients requiring subacute rehabilitation
Program components	Home care, care coordination, patient and family engagement, shared decision making, health information technology
Target population	Acute/palliative/observation/hospital-averse HaH beneficiaries: Adults who present to Mount Sinai and targeted outpatient settings, meet the MCG admission criteria for their conditions, live at home in Manhattan, and can be safely cared for at home Subacute rehabilitation (RaH) beneficiaries: Adults who are discharged from Mount Sinai, referred for RaH based on clinical and rehabilitation needs, live at home in Manhattan, and can be safely cared for at home
Theory of change/theory of action	Services for both the HaH and RaH arms may include radiology; lab work; nursing; durable medical equipment (DME); pharmacy and infusion services; telemedicine; and physical, speech, and occupational therapy. The provision of home care was intended to increase participant satisfaction and result in lower costs, superior processes, and better clinical health outcomes.
Payment model	Medicare FFS: Value-based payments, shared risk, bundled payment Other payers: Bundled payment
Award amount	$9,610,517
Effective launch date	11/18/2014
Program setting	Participants' homes
Market area	Urban
Market location	Manhattan, NY

Table I.1 *(continued)*

Program characteristic	Description
Target outcomes	**Clinical outcomes** • Decrease complications of care (for example, reduce the rate of falls) • Shorten the length of stay • Improve care process measures • Decrease the mortality rate (excluding patients receiving palliative care at home) **Cost and resource use** • Decrease 30-day unplanned readmissions • Decrease all-cause hospitalizations • Decrease total Medicare Part A and B payments • Decrease rate of hospital ED visits **Care experience** • Increase patient satisfaction

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we conclude that Mount Sinai was successful implementing the MACT program by the end of the three-year cooperative agreement. We based this conclusion on five factors–enrollment, program delivery, staffing, participant engagement, and provider recruitment and engagement. Although the awardee only enrolled 605 participants—56 percent of its enrollment target—Mount Sinai was able to implement the MACT program by the end of the cooperative agreement. In fact, Mount Sinai expanded the HaH platform to provide RaH services to adults and HaH services to children. Mount Sinai delivered these program services as intended and generally on schedule. Overall the awardee recruited, hired, and retained staff throughout the cooperative agreement, however there were two periods during which the MACT team experienced staffing shortages. While staff received trainings on the MACT program, they noted they would have preferred to receive additional trainings. Patients and caregivers participated actively in the HaH and RaH programs, engaging with MACT staff for the duration of their care. MACT clinicians reported the program had a positive effect on the access to and quality of care they provided to participants, as well as participants' satisfaction and quality of life. For the MACT program, provider recruitment and engagement was not assessed separately from program staffing and service delivery.[7]

Impact evaluation. Although the MACT intervention was implemented and even expanded, we do not anticipate being able to conduct a rigorous impact analysis for Mount Sinai because the number of beneficiaries enrolled in the treatment group was too small to ensure adequate statistical power to detect program impacts. Moreover, there were significant challenges to constructing a credible comparison group because the eligibility criteria rely heavily on clinical assessments that cannot be replicated by using Medicare FFS enrollment and claims data. However, the impact evaluation section in this narrative does include a description of the baseline characteristics of program participants.

[7] Unlike other HCIA R2 awardees, Mount Sinai's program enrolled and provided services to patients. Mount Sinai hired physicians to provide MACT services. In this report, we consider the physician recruitment part of the MACT staffing and physician engagement and a factor in effective MACT program service delivery.

Payment model. During the final year of the cooperative agreement, Mount Sinai overcame earlier challenges related to its payment model and developed models that provide bundled payments for MACT services. Mount Sinai executed a contract with a payer to establish a bundled payment for provision of acute care MACT services to the payer's Medicare Advantage and managed Medicaid beneficiaries. In addition, Mount Sinai developed a joint venture with Contessa Health to develop payment models, negotiate contracts, and process claims with other insurance companies.[8]

Sustainability plans. Mount Sinai's sustainability plans were focused on the development and execution of payment contracts with payers. However, Mount Sinai also identified opportunities to continue to provide MACT services while the contracts are negotiated. These opportunities included activities such as piloting the MACT program at a Veterans Administration medical center and providing MACT services to homebound Medicare FFS patients using house call provider codes.

[8] Contessa Health is a company that operates patient-centered prospective bundled payment programs for clinical care in patients' homes by partnering with payers and providers (http://contessahealth.com/).

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded at the start of the third program year with a sample of 6 potential respondents and achieved a response rate of 83 percent. The clinician survey was fielded from March to June of 2017 with a sample of 40 potential respondents and achieved a response rate of 67 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

During the three-year cooperative agreement, MACT enrollment strategies evolved to align with Mount Sinai's decisions to expand MACT services to increase enrollment. Potentially eligible patients for the HaH arm were identified by MACT providers in the ED, by MACT administrative assistants who reviewed the Mount Sinai electronic medical record (EMR), or through referrals. Once a potentially eligible patient was identified or referred, the MACT physician reviewed the patient's clinical information to assess whether the patient was appropriate for the program and discussed the MACT program as an option with the patient's

care team.[9] If the providers agreed that the patient would be a good candidate for the HaH arm, the MACT physician discussed the program with the patient and his or her caregivers. During these discussions, the clinician determined whether it was safe to send the patient home. If the patient agreed to participate and had a safe home environment, the patient was enrolled into the HaH arm. The MACT staff then began preparations for sending the patient home from the ED.

The enrollment process for RaH patients was similar, though they were referred to the MACT program by a caseworker, a social worker, or other health care professional.[10] During the third year of the cooperative agreement, Mount Sinai developed a protocol for reviewing appropriateness of RaH referrals. This protocol called for the MACT physician, head nurse, social worker, and therapist to review a patient's case for medical, social, and rehabilitation appropriateness prior to enrolling the patient into the MACT program.

b. Evidence of enrollment effectiveness

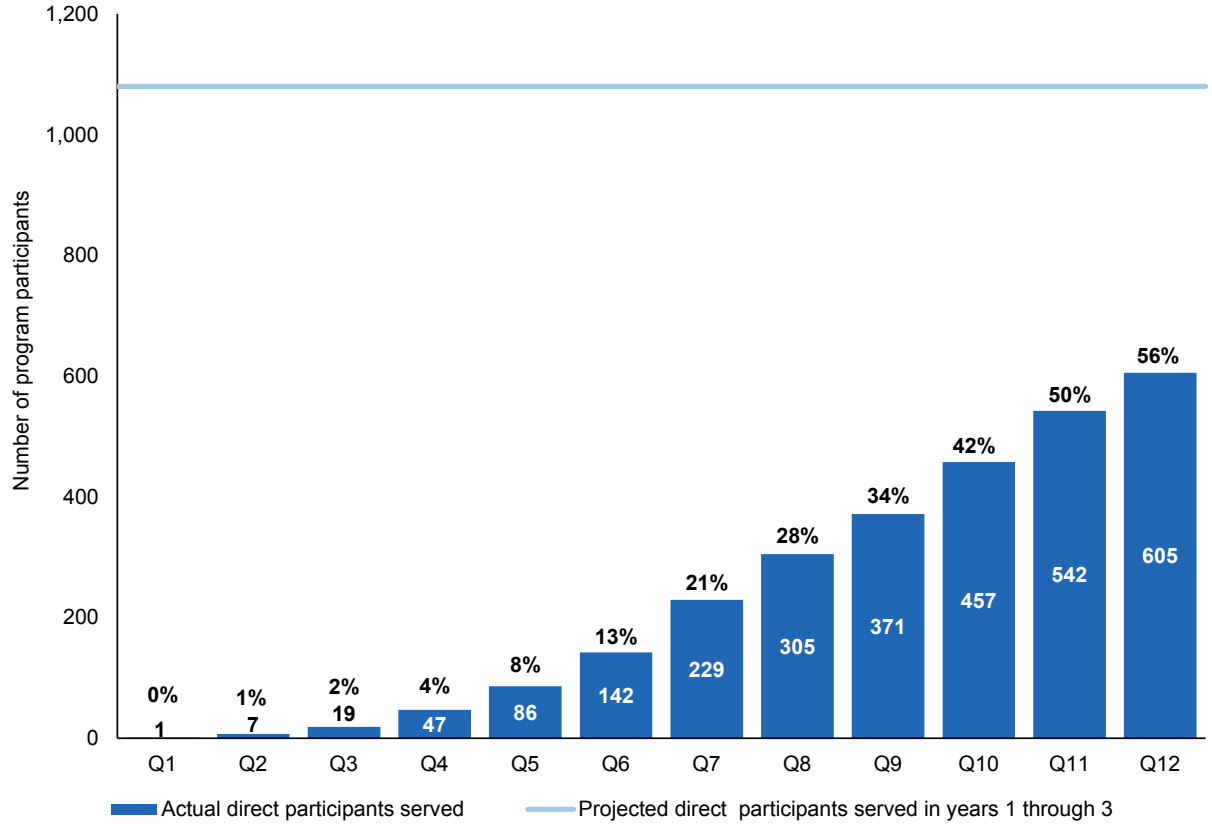
Mount Sinai was partially effective at achieving its enrollment target. Mount Sinai reported that it enrolled 605 cumulative unique direct participants from November 2014 (program launch) through August 2017.[11] This represents about 56 percent of the final three-year projections (Figure II.1). Mount Sinai did not change its target enrollment numbers during the cooperative agreement. Mount Sinai expected enrollment levels to incrementally increase as it ramped up to full capacity. Although Mount Sinai's monthly enrollment numbers increased during the cooperative agreement and it met some of its monthly enrollment targets during summer 2016, Mount Sinai had only enrolled a little over half of its target by the end of the cooperative agreement.

[9] The MACT physician reviews HaH referrals from the hospital setting (for example, emergency department or observation unit) and a MACT nurse practitioner reviews HaH referrals from the home setting.

[10] MACT administrative assistants do not review the EMR for potential RaH participants. Another difference between the HaH and RaH enrollment processes is that RaH patients are primarily discharged from their inpatient floor into the MACT program's RaH arm rather than from the ED.

[11] Unique direct participants served reflects the individuals whom the awardee served in the program through August 2017, directly paid for by HCIA R2 funding. Through August 2017, Mount Sinai had 647 cumulative total individuals ever enrolled, which also includes patients who disenrolled or opted out.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding.

c. Barriers and facilitators associated with enrollment effectiveness

Two key challenges influenced Mount Sinai's progress in meeting its three-year enrollment goals: (1) difficulty obtaining referrals to the program, and (2) a lower MACT acceptance rate than was expected Mount Sinai addressed each of these challenges by implementing changes to the eligibility criteria and recruitment processes.

Difficulty obtaining referrals. Mount Sinai found it challenging to gather sufficient referrals throughout the cooperative agreement. Several clinicians who referred patients to the program indicated in our survey that it was challenging to determine whether a patient met eligibility criteria. In addition, Mount Sinai Hospital constructed a stand-alone, 20-bed observation unit at the beginning of the cooperative agreement. This increased the hospital's capacity to provide short-stay services to patients, which may have also made it more challenging to obtain referrals to the MACT program. The MACT team tried to address the low numbers of referrals in five key ways.

First, Mount Sinai expanded the target population eligible for MACT services. Mount Sinai expanded the HaH arm's inclusion criteria and liberalized its exclusion criteria. For example, they removed the exclusion of an HIV diagnosis in Year 1 and expanded the inclusion criteria to include a dehydration diagnosis in Year 3. Mount Sinai also conducted pilot programs to understand how to develop additional program arms (and thereby expand recruitment). In the final months of the cooperative agreement, MACT staff implemented a pediatric at home pilot program.

Second, Mount Sinai shared information about the MACT program with Mount Sinai case managers and social workers, who were responsible for patients' dispositions. Therefore, they were ideal hospital staff to understand that the MACT program could be an alternative to hospital and skilled nursing facility (SNF) admissions, and share this information with treating clinicians. To increase the number of appropriate referrals for the RaH arm, one of the Mount Sinai physical therapists educated inpatient therapy staff on how to identify appropriate referrals by reviewing potential cases with them and discussing barriers that could reduce the patient's success in the home setting. Mount Sinai subsequently saw a reduction in inappropriate RaH referrals, such as patients unable or uninterested in tolerating the level of PT provided through the MACT program, those lacking appropriate home assistance to ensure their safety, or those with medical conditions making it unsafe to join RaH.

Third, Mount Sinai hired physician assistants (PAs) who already worked in the ED to also moonlight for the MACT program when they were not working their regular shifts. This was an effort to boost enrollment by expanding recruitment hours and leveraging pre-established relationships with ED doctors.

Fourth, Mount Sinai conducted a pilot to learn whether expanding recruitment to weekends would sufficiently increase enrollment. MACT staff learned how they would need to adapt enrollment processes on weekends, such as the occasional need to use a different insurance verification system, and that admissions are more likely on Saturdays than Sundays. While Mount Sinai gained insight from this pilot, MACT staff determined that unless they executed payment contracts with additional payers and increased the pool of eligible participants, it would not be cost-effective to offer MACT weekend admissions.

Finally, in the third year of the cooperative agreement, Mount Sinai expanded the HaH arm to other Mount Sinai hospitals, including St. Luke's, Beth Israel, and Mount Sinai West to increase the number of referrals. However, the MACT team found that just expanding the program to additional locations did not improve recruitment. They learned it also takes local buy-in for hospital staff to generate referrals, which needs to be developed through: (a) continuous relationship building (for example, by attending internal medicine meetings to explain the MACT program and present on patients eligible for MACT from that hospital's ED, meeting with case managers to explain the program and provide referral criteria), (b) MACT staff connecting physicians at the new hospital with physicians at the main Mount Sinai hospital who support the program and are enthusiastic about referring potential participants, and (c) providing financial support to formal program champions.

Lower-than-expected program acceptance rate. Mount Sinai had expected around 90 percent of eligible participants to enroll in the MACT program, as this was the acceptance rate

documented by other HaH programs. Initially, Mount Sinai reported a MACT acceptance rate closer to 60 percent. However, during the last year of the cooperative agreement, Mount Sinai reported a monthly acceptance rate of approximately 86 percent. In the survey, clinicians and staff described barriers that may have contributed to patient refusals, including too much of a time commitment and too many requirements, being happy with the old system of care, and concerns around privacy/confidentiality.

2. Delivery of program services

a. Description of and changes to service delivery model

Originally, Mount Sinai planned to use the HCIA R2 funding to provide acute care services to patients in their home in hopes of reducing potential adverse consequences of hospitalization (such as pressure ulcers, falls, and delirium). Over the course of the cooperative agreement, Mount Sinai expanded the HaH arm of the MACT program to also provide palliative care and care to patients who are hospital averse and created the RaH arm to provide subacute rehabilitation care to patients in their homes.

During the acute care phase of the HaH arm, the team provided acute care services, such as daily physician or nurse practitioner visits, one to two daily nurse visits, and additional services depending on the patient's clinical and social needs. In the post-acute care phase, the team provided the patient with necessary transitional services, such as physician and nurse visits, outpatient care coordination, and social work services, to prevent hospital readmissions. During the subacute rehabilitation phase of the RaH arm, the team provided daily PT that typically lasted for 45 minutes to one hour, six 6 days per week and, as needed, clinical services, such as nurses' visits, pharmacy or infusion services, or additional therapy services, such as occupational or speech therapy. In the post-subacute rehabilitation phase of RaH, the team provided transitional services such as outpatient care coordination, and social work services.

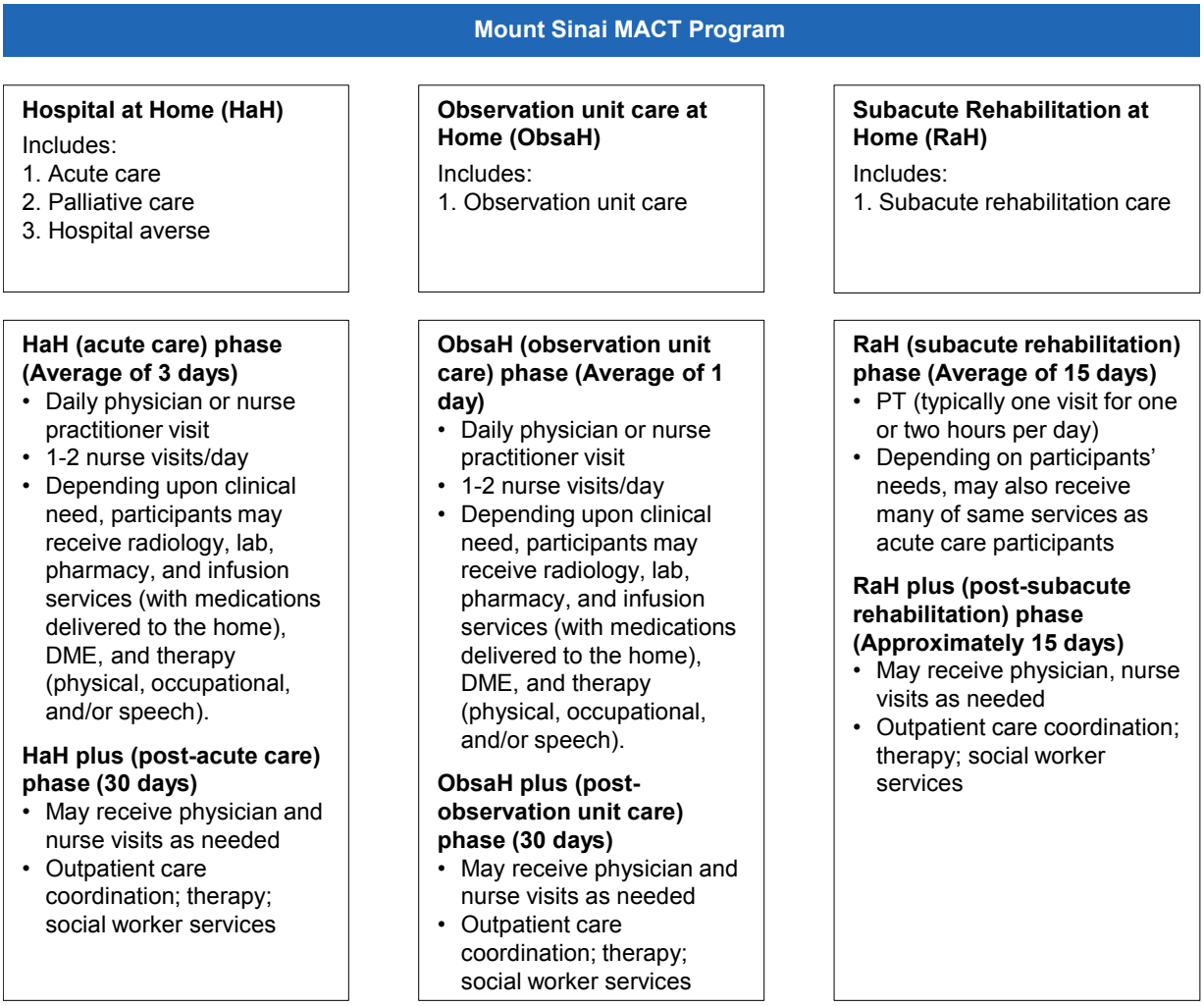
Infrequently, a MACT HaH participant's clinical condition worsened during his or her MACT episode. Mount Sinai categorized these situations into two categories based on when the patient's deterioration occurred. If the deterioration occurred during the acute care phase of the MACT program and the patient had to return to the hospital to receive the appropriate level of care, Mount Sinai considered this to be an escalation of care. If the deterioration occurred after the patient had been discharged from the acute care phase of the HaH arm and the acute care phase either had to be resumed, or the patient had to be admitted to the hospital, Mount Sinai considered this to be a readmission.

Initially, Mount Sinai relied on external contractors to provide MACT nursing and therapy services. As the program evolved, Mount Sinai reduced its reliance on external contractors by hiring Mount-Sinai employed nurses and physical therapists. These changes resulted in the core MACT team being comprised of only Mount Sinai staff in the final year of the cooperative agreement. One benefit of this restructuring was that, unlike in previous years, in which contracted staff might vary from day to day and might attend team meetings or not, this consistent group of Mount Sinai staff huddled every day and shared information on patients' status (including clinical, social, and therapy aspects of care).

During the third year of the cooperative agreement, Mount Sinai hired several new staff, including a physician, two nurses, and two physical therapists. Mount Sinai hired a physician full time. This was a first for the MACT program, as all other MACT physicians had worked on a rotating, part-time basis. Mount Sinai hired one full time and one per-diem nurse to expand the program's patient capacity. Mount Sinai hired the two physical therapists to oversee the rehabilitation aspect of the MACT program and provide therapy services. All trainings of new MACT staff focused on teaching about the MACT program and the staff members' roles in the program. The trainings were either done by the employees' supervisors or were completed by the MACT leadership team.

While Mount Sinai began establishing MACT program procedures (for example, workflow processes) during previous years of the cooperative agreement, protocol development was a priority in the final year as well as completion of an implementation guide for interested health systems as part of the MACT dissemination strategy. Protocols were developed through round table discussions—a MACT physician drafted a protocol and sent it out to the entire MACT group, after which the team discussed and revised the protocol. The MACT physician used this team-based approach to foster MACT staff acceptance of the protocols. However, team members acknowledged that additional revisions would be made to the protocols to improve program processes. The team also had flexibility regarding protocol adherence based on specifics of patients' circumstances. The team began establishing protocols regarding care escalation processes, social work assessment timing, and how to provide MACT care to pediatric patients. The protocols explained what the program can and cannot do, how to assign roles to certain team members, and how to identify available resources.

Figure II.2. MACT: Hospital at Home and Rehabilitation at Home Program Arms



Note: This figure represents the MACT program as it was functioning at the end of year 3. Prior to year 3, ObsaH was considered a component of the HaH arm.

[a]The HaH observation unit care phase lasts one day on average.

b. Evidence of service delivery effectiveness

Mount Sinai achieved service delivery effectiveness and implemented the proposed HaH program on schedule. Overall the awardee recruited, hired, and retained staff throughout the cooperative agreement, however staff noted a desire for additional trainings. Mount Sinai also created and implemented the RaH program arm and patients participated actively in the HaH and RaH programs, engaging with MACT staff for the duration of their care. MACT clinicians reported the program had a positive effect on the access to and quality of care they provided to participants, as well as participants' satisfaction and quality of life. Since, unlike some other HCIA R2 awardees, the MACT program did not have to recruit and engage providers, as all

MACT services were rendered by employed or contracted providers, we did not perform a separate assessment of provider recruitment and engagement.

Delivery of intervention services. Mount Sinai not only provided the acute and post-acute services in the patient's home as proposed, it also expanded to provide subacute rehabilitation services. The frequency, intensity, and duration of MACT services provided were matched with the patient's clinical needs. For example, if patients had complex wounds or needed infusions with long durations, MACT nurses would see fewer patients that day and spend the necessary additional time with those who needed more care.

Staffing and training. Mount Sinai retained almost all MACT program staff and most of those who left did so due to reasons unrelated to the program. MACT team members expressed admiration of and confidence in their MACT colleagues. Although the majority of the MACT staff participated in formal trainings, a third of the clinicians who provided direct MACT patient care and responded to the clinician survey wished that additional trainings had been offered to help them meet their MACT program responsibilities. At times the program team felt the stress of staffing shortages, including the final program year during which the team experienced physician shortages. For example, 13 percent of the clinicians who provided direct MACT patient care and responded to the clinician survey thought their overall load for the MACT program was somewhat heavy. Forty percent of these clinicians reported the MACT program increased their feelings of burnout at work a little or a lot. However, all staff completing the non-clinician survey reported that their overall load was about right, indicating that this was more of an issue for clinicians.

Recruitment and engagement of providers. This domain was not applicable to the MACT program because Mount Sinai enrolled patients, not providers, into the MACT program. Mount Sinai hired providers to provide MACT services and the engagement of these providers contributed to high quality MACT services delivery, which is discussed in the service delivery section.

Engagement of program participants. The majority of patients who enrolled in the MACT program were engaged for the duration of their care. Patients and their caregivers in both MACT arms were required to be engaged in the patients' care because many of the activities, such as helping the patient to the bathroom or preparing food, which would have been completed by nurses or support staff in an inpatient setting, were left to the patient and his or her support system to complete. The majority of MACT clinicians, who provided direct patient care and responded to the clinician survey, strongly agreed that the MACT team had successfully engaged patients with the program (87 percent).

c. Barriers and facilitators associated with service delivery effectiveness

Delivery of intervention services. The MACT staff's motivation to improve care processes and the team's flexibility and creativity fostered continual program adaptations and effective delivery of the MACT program.

Continuous process improvement. MACT staff improved program processes throughout the cooperative agreement, which facilitated program implementation. For example, MACT staff completed a social work workflow assessment to learn how different processes would affect

patient care and worker satisfaction. One change made based on the results of this assessment was that social workers assigned to MACT participants started introducing themselves to patients in the hospital before, rather than after, the patient had been discharged to home and admitted into the MACT program. This initial introduction helped reduce the surprise of the social worker showing up at the patient's home for the initial visit. Mount Sinai's commitment to improve MACT program processes was also illustrated in the team's continued efforts to use its evolving knowledge of the MACT program to enhance the EMR system's documentation tools.

Flexibility, engagement, and creativity of MACT staff. The MACT staff's flexibility, engagement, and creativity were major facilitators that allowed them to effectively implement the MACT program. When faced with challenges, Mount Sinai staff identified potential solutions and implemented them even if the changes meant diverting from their initial plan. For example, Mount Sinai originally designed the MACT program to be primarily based on outsourced services, in order to demonstrate that the program could be implemented by a variety of organizations as long as contracts had been established with other organizations to provide services that were not available from the primary organization. Mount Sinai contracted with another organization to provide nursing and PT services to MACT patients because Mount Sinai did not have nursing staff to provide home infusion and other home-based skilled nursing services or therapy staff that provided home-based PT. However, as the program progressed, Mount Sinai felt the MACT program would benefit from more consistent nurses and physical therapists providing MACT services and a higher level of communication between the nurses and physical therapists and the rest of the MACT staff. After failed attempts at resolving these concerns with the contracted organizations and one vendor's decision to no longer provide the specialized nursing service that the MACT program required, Mount Sinai decided to hire nurses and physical therapists internally to provide MACT services. MACT staff reported that hiring these two physical therapists improved staff communication and service delivery.

Engagement of program participants. MACT physical therapists motivated patients to actively participate in the program by drawing on their desire for independence.

Focus on independence. The Mount Sinai physical therapists strived to make the program as engaging as possible for participants. The physical therapists used activities, such as going to the store, as part of their therapy to spark patients' desire for independence as a motivating factor. In addition, unlike therapy provided by other programs or in other facility settings, MACT physical therapists gave patients a narrow window in which to expect the PT to arrive and provide services so that patients could plan their day, live as independently as possible, and be active, which compliments progress made in their therapy sessions.

Staffing and training. Strong team cohesion and a flexible staffing structure facilitated program implementation, while staffing shortages and the need for more robust training were barriers for the MACT program.

Team cohesion. The MACT staff had strong team cohesion that was built on mutual support and respect, which facilitated program implementation. This may have been one reason why Mount Sinai did not have difficulty retaining providers during the three-year cooperative agreement. MACT team members who left the team did so for reasons external to the program, such as the provider completing his or her residency.

> "Staff in the MACT program are happy here…they share the same passion. Working with the team provides support, so that even in tough times you're not alone and someone has your back. We're like a family here. We celebrate every birthday, every baby shower, every holiday – so we go through stress together, but we recover together"
>
> —MACT staff member

Staffing structure. The MACT staff's continual modification of the staffing structure to identify the most effective and efficient staff organization was a program facilitator. For example, in the final year of the cooperative agreement, Mount Sinai formalized the nurse manager position, which had developed through the evolution of the MACT program. As program enrollment increased, MACT staff needed someone to coordinate the program's activities, which resulted in one nurse practitioner staying in the office to complete tasks such as assessing program capacity to take on additional patients, triaging issues identified by MACT staff in the field, diverting resources to the patients with the highest need, and overseeing logistical activities. The combination of team cohesion and enhanced communication produced efficient, seamless care. For example, if the PT was visiting a MACT RaH patient and the patient's clinical status was noticeably worse than the previous day (for example, confused, lethargic), the PT could call the MACT nurse manager, who could reroute a nurse to visit the patient quickly and reschedule the nurse's next visit for later in the day. Mount Sinai also expanded the MACT staff's capacity by more effectively delegating activities to administrative assistants. In the third year of the program, the administrative assistants began taking responsibility for activities such as scheduling follow-up appointments, notifying patients of appointments, and keeping track of supplies in the homes of patients and coordinating their return as feasible. The project coordinator oversaw the administrative assistants. The goal of this delegation was to free up the clinical staff to focus on other clinical activities. Surveyed clinicians had described insufficient program staff time for the amount of work required as either a major barrier (7 percent) or a minor barrier (20 percent).

Staffing shortages were a barrier, albeit intermittent and temporary, to service delivery implementation. Although the MACT program successfully retained the majority of its staff over the three-year cooperative agreement, Mount Sinai did experience MACT staffing shortages at two points during the cooperative agreement. The first staffing shortage occurred in early 2016. As MACT enrollment began increasing, MACT staff noted that the clinical acuity level of the MACT caseload could vastly affect the level of staffing needed to provide MACT services (for example, sometimes MACT nurses could care for a maximum of three patients a day because their patients had complex wounds or long infusions; other days nurses could care for six patients a day because their patients needed less time-consuming services). While trying to identify a staffing model that could appropriately handle such variable staffing needs, Mount Sinai requested and received a technical assistance site visit by other HCIA R2 contractors, during which the contractors provided feedback on Mount Sinai's staffing model. Mount Sinai decided to hire per diem nurses and contract with an infusion nursing agency to provide additional staff when needed. In addition, as described below, Mount Sinai began using one of

the MACT NPs to better manage staffing capacity and take overall responsibility for scheduling and supervising nursing staff.

The second staff shortage Mount Sinai experienced was towards the end of the cooperative agreement when three MACT providers moved out of state. As a result, the team was left with a solo physician responsible for recruitment from all four Mount Sinai hospitals (Mount Sinai, St. Luke's, Beth Israel, and Mount Sinai West) at a time when two of the hospitals were just starting to recruit for the MACT program. While managing recruitment during the physician shortage was difficult, Mount Sinai dealt with this staff shortage by hiring one physician and one nurse practitioner.

Training. Although survey results indicated that 80 percent of MACT staff received formal training, the program struggled to provide training in a way that fully met staff needs. Over a third of the staff who provided MACT services indicated that they would have benefited from additional trainings more directly related to their responsibilities. MACT staff developed process protocols during the third year of the cooperative agreement to offer additional training support. The team felt these protocols could aid in training activities for future staff and improve the consistency of program implementation processes across current staff. The protocols could also be used for program replication efforts in other locations as part of the sustainability efforts.

C. Assessment of perceived program effects on the delivery of care and outcomes

The majority of staff who provided direct MACT care and responded to the non-clinician or clinician survey perceived that the MACT program had positive impacts on the delivery of care and patient outcomes (Table II.2). For example, the majority of staff thought the MACT program had been very effective (85 percent) or somewhat effective (10 percent) in achieving its goals. The majority of them also indicated that the program had a positive impact on the quality of care and services they provided to patients (95 percent), patients' satisfaction (90 percent), quality of life (90 percent), and access to care (100 percent), as well as achievement of participants' health goals (95 percent). Most of them strongly agreed that the MACT program was making a difference in meeting critical needs in the community (84 percent).

> "I have seen this program grow both in patient quantities and staff quantities. I have witnessed the transformation of certain patients after our care and the family members are still happy till this day. From someone who has watched her friend deteriorate due to multiple hospitalizations, I wish this would have been around for her."
>
> —MACT staff member

Several MACT respondents also believed the program would result in positive patient outcomes for several clinical reasons related to treating patients in their homes. For example, patients in the MACT program are not treated in traditional hospitals in close proximity to other patients with infections. In addition, MACT program providers gain additional context about patients' living conditions, which can help identify solutions to clinical issues. However, some interview respondents also acknowledged that the small number of participants may pose a challenge to detecting program impacts.

Table II.2. Staff perceptions of MACT program effects on care

Percent of staff providing direct care indicating that the MACT program had a positive impact on the following:	Percentage of respondents
The quality of care and services you provide to participants	95
Your ability to respond in a timely way to participant needs	90
Your ability to provide care or services that are responsive to participant preferences, needs, and values[a]	100
Access to care or services for all participants	100
Achievement of participants' health goals	95
Participant satisfaction	90
Participant quality of life	90
Care coordination	95

Source: HCIA R2 evaluation survey of participating non-clinician and clinician staff, fall 2016. Clinicians who only referred patients to the program were excluded.

Note: N=20 for all items except as noted

[a]One missing response.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Throughout the three-year cooperative agreement, Mount Sinai's goal was to implement a HaH program that provided high quality care to several target populations. However, without a wealth of clinical and social data, the targeting criteria themselves would be difficult to reproduce in the construction of potential comparison groups in an impact analysis. It would be impossible to duplicate in any available data, for example, the physician, nursing, social worker, and therapist approvals required for MACT enrollment (for instance, therapists' assessments of the ability to tolerate PT); to determine whether a patients' homes had a shared bathroom or not; or whether patients had sufficient family support to help with daily activities (such as meal preparation or telephone use). Certainly, Medicare FFS claims and enrollment data lack this level of detail. The program team's flexibility with modifying inclusion and exclusion criteria to select patients who would be a good fit for the program rather than sticking with strict inclusion and exclusion criteria, while a strength of the program, would further complicate any efforts to select a comparison group for an impact analysis—for example, any potential comparison group members could have HIV diagnoses in Years 2 and 3 but not Year 1, and diagnoses of dehydration in Year 3 but not in Years 1 and 2. In general, it will be difficult if not impossible to conduct a non-random assignment, retrospective impact analysis on any program that relies heavily on clinician judgment and unmeasured clinical and social factors to select participants.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Mount Sinai's programs and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Tables III.1a and III.1b: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1a. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: Mount Sinai Hospital at Home

Evaluability domain	Response
Projected number of Medicare FFS population with 6 months of program exposure by February 28, 2018	184[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,220
Likelihood of all-cause hospitalizations	553
MDE sample size requirement to detect 20% effect	
Total expenditures	305
Likelihood of all-cause hospitalizations	138
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	No, low rate of identification of treatment group with claims data with significant dilution of treatment effect within an intent-to-treat framework
Likelihood of solid comparison group	Serious concern. We may be not able to identify a strong comparison group.
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Too few treatment beneficiaries for primary outcomes
Survey data for treatment group that will be analyzed	Clinician, staff, and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

FFS = fee-for-service.

Table III.1b. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: Mount Sinai Rehabilitation at Home

Evaluability domain	Response
Projected number of Medicare FFS population with 6 months of program exposure by February 28, 2018	230[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	969
Likelihood of all-cause hospitalizations	763
MDE sample size requirement to detect 20% effect	
Total expenditures	242
Likelihood of all-cause hospitalizations	191
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	No, low rate of identification of treatment group with claims data with significant dilution of treatment effect within an intent-to-treat framework
Likelihood of solid comparison group	Serious concern. We may be not able to identify a strong comparison group.
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Too few treatment beneficiaries for primary outcomes
Survey data for treatment group that will be analyzed	Clinician, staff, and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

FFS = fee-for-service.

The HaH program, in our view, deserves intensive future study. We find, however, that three features of the program limit its evaluability under HCIA R2:

1. **There is no means of selecting an appropriate comparison group.** In order to increase recruitment into the program for HCIA R2, Mount Sinai relaxed its highly specific inclusion criteria to permit greater physician judgment in determining which patients could safely be treated in the home. Even with access to EMR data, replicating the judgment of Mount Sinai physicians to select a comparison group would be unrealistic. Moreover, patients could be dropped from the HaH program if the home environment was considered unsuitable or unsafe. There is no realistic means of assessing the home environment for patients in a potential comparison group. Although it is certainly possible to match HaH participants to nonparticipants in terms of diagnoses and prior inpatient and outpatient utilization by using

Medicare data, there is no realistic way of replicating the judgment process used to select HaH participants.[12]

2. **The control group selected by Mount Sinai Hospital may not be well matched to HaH participants.** Mount Sinai describes its control group as consisting of "those patients eligible for MACT [mobile acute care teams] HaH but who present to the emergency department at times when the MACT program does not evaluate new patients (evenings, weekends)." Whether patients who arrive at the emergency department in the evening or on weekends resemble those who arrive when the MACT teams are working is unclear. We are currently investigating whether or not these patients received the same MACT evaluation process received by HaH participants. At this point, it remains unclear whether the control group can be considered an appropriate counterfactual for estimating impacts.

3. **The sample of HaH patients available for analysis is too small to detect program impacts, even if those impacts are sizable.** The final file for the HaH program submitted by Mount Sinai Hospital contains 184 Medicare FFS beneficiaries enrolled in both Part A and Part B of Medicare. (The Mount Sinai–selected control group contains 104 beneficiaries.) Our analysis indicates that this sample would be insufficient to detect an impact on Medicare expenditure as large as 20 percent.

B. Characteristics of Medicare and Medicaid participants at baseline

This section summarizes the common and awardee-specific claims-based outcomes at baseline for the treatment group, which we measured for the 12 months before each beneficiary's enrollment date in MACT. At the end of August 2017, the program had 682 participants, including 420 participants for the four subgroups of the HaH arm ("purely" acute, 297; observation unit, 64; palliative care at home, 26; hospital averse at home, 33) and 262 participants for the RaH arm.[13] Given the small sample sizes for three of the four HaH subgroups, for the purpose of this evaluation, we focused on the "purely" acute population for the HaH arm and the RaH population for the RaH arm. Because the need for medical care and the services provided in each arm are significantly different from one another, we evaluated each arm separately.

[12] One approach often used in situations such as this is to model the selection process by using available claims data and include in the treatment group all patients who had high predicted probabilities of being included (whether or not they actually were treated). Outcomes for this group could then be compared to outcomes for a comparison group of patients with similar characteristics and predicted probabilities of inclusion who were treated at a different hospital in New York City. This approach, however, would make detecting a meaningful difference in outcomes even more difficult. (See discussion of sample size limitations).

[13] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

RaH arm. In presenting the baseline characteristics, we restricted the treatment group to adult beneficiaries (age 18 and older) in Medicare fee-for-service (FFS), Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they must have been enrolled in the awardee's program on or before August 31, 2017, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. Each individual's enrollment date is determined by the start of Mount Sinai's RaH care at home. The calendar period covered by the baseline quarters is determined by the enrollment date for each participant and therefore varies by participant. After we excluded beneficiaries who did not meet the above criteria, 230 participants were included in the analysis of baseline characteristics for this report.

The Medicare FFS beneficiaries participating in the RaH arm of the MACT program were a predominantly elderly group with significant, high care needs (Table III.2a). More than half of the recruited beneficiaries were older than 85 (57 percent); only 3 percent were younger than 65. A majority of participants were originally eligible for Medicare based on age (87 percent), while 12 percent were eligible because of a disability. In addition, 17 percent were dually eligible for Medicare and Medicaid, comparable to the national level (18 percent). Participants were far more likely to be female (67 percent) and white (73 percent). Overall, Mount Sinai was recruiting a population in poor health whose Medicare expenditures would likely be high in the future, as evidenced by the fact that the average hierarchical condition category (HCC) risk score for participants (3.5) was more than triple the average score for Medicare FFS beneficiaries nationwide (1.00).

Consistent with their high needs, participants in the RaH arm had high rates of Medicare expenditures and service use in the year prior to enrollment, particularly in the last quarter. Table III.3a shows baseline utilization and expenditure data for a common set of measures, including the four CMMI core measures. We examined the baseline costliness of care by calculating average per beneficiary per month (PBPM) Medicare payments in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $5,571. The average PBPM Medicare payments for inpatient care ($3,203), physician services ($817), and SNF ($509) were the largest drivers of the total cost of care, representing 81 percent of this cost. Through the RaH arm, Mount Sinai expects to reduce expenditures on high-cost services in the SNF setting relative to what would have occurred absent the program by offering convenient RaH services at home. If reductions in the utilization of these high-cost services are realized, then the awardee should see a similar decline in Medicare expenditures.

The average utilization of expensive Medicare services before enrollment was also high, particularly in the last quarter. Nearly all participants (94 percent) were hospitalized, resulting in an annual rate of acute care hospitalizations of 2,043 per 1,000 participants during the baseline year.[14] ED utilization was also high. Fifty-three percent of participants visited an ED in the baseline year, leading to an annual rate of outpatient ED visits of 975 per 1,000 participants. Similarly, the 30-day unplanned readmission rate for participants was 32 percent per discharge,

[14] The three-day inpatient hospital stay required for a covered SNF stay does not apply to the RaH arm, so we are not expecting to see a 100 percent hospital admission rate during the baseline year.

or 16 percent per participant, which is much higher than the national rate of 18 percent per discharge. These findings indicate that there may be a sizeable opportunity to reduce potentially avoidable admissions, ED visits, and readmissions during the intervention period by providing effective RaH services at home. At the same time, participants had high rates of primary care utilization (12,292 primary care visits per 1,000 participants per year) and specialty care utilization (33,073 specialist visits per 1,000 participants per year).

There was a dramatic rise in expenditures and utilization in the last quarter before enrollment, corresponding to the high probability of hospitalization. As shown in Table III.3a, the total average PBPM Medicare payment rose from 2,644 in the first quarter of the baseline year to 12,906 in the last quarter before enrollment.

HaH arm. In presenting the baseline characteristics, we restricted the treatment group to adult (age 18 and older) beneficiaries in Medicare FFS, Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they must have been enrolled in the awardee's program on or before Aug 31, 2017, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. Each individual's enrollment date is determined by the start of Mount Sinai's HaH care at home. The calendar period covered by the baseline quarters is determined by the enrollment date for each participant and therefore varies by participant. After we excluded beneficiaries who did not meet the above criteria, 184 participants were included in the analysis of baseline characteristics for this report.

The Medicare FFS beneficiaries participating in the HaH arm of the MACT program were a predominantly elderly group (Table III.2b). More than three-quarters of the recruited beneficiaries were 75 years or older (78 percent). Most participants were originally eligible for Medicare based on age (84 percent), while 15 percent were eligible because of a disability. In addition, 30 percent were dually eligible for Medicare and Medicaid. This indicates a high level of social need, considering that 18 percent of beneficiaries nationwide are dually eligible. Participants were far more likely to be female (72 percent) and white (67 percent). Overall, participants in the HaH arm were substantially less healthy and had a greater need for care than the general Medicare FFS population, as evidenced by the fact that the average HCC risk score for participants (2.8) was almost triple the average score for Medicare FFS beneficiaries nationwide (1.0).

Consistent with their high needs, participants had high rates of Medicare expenditures and service use in the year prior to enrollment. Table III.3b shows baseline utilization and expenditure data for a common set of measures, including the four CMMI core measures. We examined the baseline costliness of care by calculating average PBPM Medicare payments in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $3,167, which is substantially higher than the 2014 national average of $792.[15]

[15] The national data here and in the next paragraph are from the CMS, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

The average PBPM Medicare payment for inpatient care ($1,521) was the largest driver of the total cost of care, representing almost half of this cost (48 percent). Through its HaH arm, Mount Sinai intends to reduce the total costs to CMS by more than 50 percent for the 30-day episode of care (when compared with the 30-day cost of care for an individual with similar health problems who receives care through a traditional inpatient hospitalization).[16] If reductions in the costs for the 30-day episode of care are realized, then the awardee should see a similar decline in Medicare expenditures.

Participants in the HaH arm had high average utilization of expensive Medicare services before enrollment. More than three fourths of the participants (78 percent) had an ED visit, resulting in an annual rate of outpatient ED visits of 1,651 per 1,000 participants during the baseline year. Fifty-eight percent of participants had a hospital admission in the baseline year, leading to an annual rate of acute hospital admissions of 1,188 per 1,000 participants. Similarly, the 30-day unplanned readmission rate for participants (23 percent) was much higher than the national rate of 18 percent per discharge. In addition, the annual rate of specialist visits (21,155 per 1,000 participants) was substantially higher than the rate of primary care visits (10,773 per 1,000 participants). These findings indicate that there may be a sizeable opportunity to reduce potentially avoidable admissions, ED visits, and readmissions during the intervention period by providing effective HaH services at home.

We observed an upward trend in average PBPM total payments in the last quarter before enrollment (approximately a 58 percent increase from the total PBPM average of quarters 1, 2, and 3). This is mainly a result of the increase in acute inpatient and outpatient care. We observed a similar pattern in the rates of hospital admissions, outpatient ED visits, and observation stays. These increases indicate that the beneficiaries targeted for the HaH arm are high-cost individuals who used acute care services extensively in the year before enrollment and, in particular, in the quarter that immediately preceded enrollment.

[16] The 50 percent reduction excludes the core MACT services that are paid for with HCIA R2 funds. This target is more easily measured than the estimated 20 percent run rate savings, which include the cost of the core MACT services.

Table III.2a. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the RaH arm of Mount Sinai's program through August 31, 2017

	All participants (N = 230)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	8	3
65 to 74	18	8
75 to 84	73	32
85 and older	131	57
Gender		
Female	153	67
Male	77	33
Race		
White	168	73
Black	39	17
American Indian, Alaska Native, Asian/Pacific Island American, or other	10	4
Hispanic	12	5
Original reason for Medicare eligibility		
Old age and survivor's insurance	200	87
Disability insurance benefits	27	12
End-stage renal disease (ESRD)[a]	3	1
Hospice[b]	1	0.43
Medicare/Medicaid dual status, percentage dual[b]	38	17
HCC score[c]		Statistic
Mean		3.52
25th percentile		1.89
Median		3.19
75th percentile		4.81

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of August 31, 2017.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the start of the beneficiary's RaH care at home. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3a. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the RaH arm of Mount Sinai's program through August 31, 2017

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	230	225	225	230	230
Average Medicare expenditures PBPM[a]					
Total	5,571 (357)	2,644 (441)	2,754 (424)	3,773 (504)	12,906 (847)
Acute inpatient	3,203 (231)	1,173 (258)	1,057 (240)	1,485 (290)	8,947 (671)
Inpatient other[b]	354 (75)	181 (136)	163 (101)	265 (139)	796 (199)
Outpatient[c]	322 (40)	195 (31)	225 (41)	350 (76)	511 (79)
Physician services	817 (46)	552 (51)	565 (52)	629 (51)	1,504 (90)
Home health	279 (28)	257 (166)	225 (36)	262 (42)	369 (40)
Skilled nursing facility	509 (89)	211 (87)	420 (127)	670 (176)	714 (133)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	2 (2)
Durable medical equipment	87 (59)	75 (49)	99 (73)	113 (86)	63 (29)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	2,043 (145)	918 (162)	891 (185)	1,222 (212)	5,043 (206)
Outpatient ED visits[e]	975 (115)	684 (129)	570 (116)	903 (152)	1,722 (221)
Primary care visits in any setting	12,292 (934)	8,080 (982)	8,837 (1,045)	9,401 (1,079)	22,574 (1,580)
Primary care visits in ambulatory settings	7,227 (576)	5,903 (611)	6,432 (588)	6,675 (640)	9,826 (945)
Specialist visits in any setting	33,073 (1,868)	22,459 (1,812)	23,731 (2,044)	26,381 (2,078)	58,730 (3,808)
Specialist visits in ambulatory settings	20,068 (1,216)	17,888 (1,367)	19,366 (1,532)	20,361 (1,537)	22,383 (1,417)

Table III.3a *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	94 (2)	17 (2)	13 (2)	22 (3)	90 (2)
Percentage with an outpatient ED visit[e]	53 (3)	14 (2)	11 (2)	18 (3)	31 (3)
Percentage with a 30-day readmission among all discharges	32 (3)	28 (7)	37 (6)	30 (6)	31 (5)
Percentage of participants with a readmission among all participants	16 (2)	2 (1)	4 (1)	4 (1)	11 (2)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of August 31, 2017.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

Table III.2b. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the HaH arm of Mount Sinai's program through August 31, 2017

	All participants (N = 184)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	13	7
65 to 74	28	15
75 to 84	50	27
85 and older	93	51
Gender		
Female	132	72
Male	52	28
Race		
White	124	67
Black	34	18
American Indian, Alaska Native, Asian/Pacific Island American, or other	5	3
Hispanic	18	10
Original reason for Medicare eligibility		
Old age and survivor's insurance	155	84
Disability insurance benefits	27	15
End-stage renal disease (ESRD)[a]	2	1
Hospice[b]		
Medicare/Medicaid dual status, percentage dual[b]	56	30
HCC score[c]		Statistic
Mean		2.75
25th percentile		1.47
Median		2.29
75th percentile		3.63

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of August 31, 2017.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the start of the beneficiary's HaH care at home. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; ESRD = end-stage renal disease; HCC = hierarchical condition category.

Table III.3b. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the HaH arm of Mount Sinai's program through August 31, 2017

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	184	180	180	183	184
Average Medicare expenditures PBPM[a]					
Total	3,167 (299)	2,118 (271)	3,257 (599)	2,854 (436)	4,346 (423)
Acute inpatient	1,521 (198)	886 (196)	1,506 (336)	1,359 (293)	2,250 (313)
Inpatient other[b]	76 (56)	0 (0)	222 (220)	38 (38)	46 (45)
Outpatient[c]	300 (33)	188 (31)	170 (27)	248 (53)	587 (72)
Physician services	532 (37)	443 (49)	520 (47)	500 (49)	663 (49)
Home health	409 (40)	344 (53)	429 (61)	356 (51)	507 (64)
Skilled nursing facility	244 (58)	164 (72)	304 (126)	280 (111)	227 (83)
Hospice	26 (18)	38 (32)	60 (42)	5 (5)	0 (0)
Durable medical equipment	59 (10)	55 (20)	46 (10)	69 (20)	66 (14)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	1,188 (138)	677 (137)	1,106 (170)	1,042 (206)	1,864 (238)
Outpatient ED visits[e]	1,651 (206)	903 (202)	1,151 (275)	842 (157)	3,662 (383)
Primary care visits in any setting	10,773 (823)	8,102 (734)	10,044 (1,045)	11,238 (1,368)	13,617 (1,163)
Primary care visits in ambulatory settings	8,108 (536)	6,838 (645)	7,855 (790)	8,157 (777)	9,538 (878)
Specialist visits in any setting	21,155 (1,387)	18,641 (1,585)	19,682 (2,052)	19,062 (1,636)	27,102 (2,240)
Specialist visits in ambulatory settings	16,159 (1,136)	15,866 (1,399)	15,596 (1,527)	14,208 (1,183)	18,923 (1,592)

Table III.3b *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	58 (4)	15 (3)	23 (3)	19 (3)	34 (4)
Percentage with an outpatient ED visit[e]	78 (3)	15 (3)	18 (3)	16 (3)	68 (3)
Percentage with a 30-day readmission among all discharges	23 (3)	18 (7)	13 (6)	28 (7)	27 (6)
Percentage of participants with a readmission among all participants	10 (2)	1 (1)	2 (1)	5 (2)	4 (2)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of August 31, 2017.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month; SNF = skilled nursing facility.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

During the final year of the cooperative agreement, Mount Sinai completed the development of payment models that provide bundled payment for MACT services. Mount Sinai executed a contract with a payer to establish a bundled payment to provide acute care MACT services to the payer's Medicare Advantage and Medicaid managed care beneficiaries. In addition, Mount Sinai developed a joint venture with Contessa Health to develop payment models, negotiate contracts, and process claims with other insurance companies.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

The basic HaH payment model was a bundled payment covering the provision of core MACT services in the acute care phase that currently do not have a payment mechanism. Mount Sinai simultaneously designed variations of this basic model to align with the needs of individual payers for patients covered under Medicare Advantage, managed Medicaid, commercial, and Medicare FFS plans. Depending on the preferences of the specific payer, the bundled payment model could be expanded to include the post-acute care phase of the HaH arm and incorporate a shared risk component. Mount Sinai was also separately constructing a payment model for the RaH arm.

For Medicare FFS, Mount Sinai developed a single payment model that would cover service provision to three of its target populations, including those receiving acute care, palliative care, and observation unit services. The bundled payment would cover core MACT services in the acute and post-acute phases of the MACT program's HaH arm (but not the RaH arm).[17] Core services covered by the bundled payment include items such as MACT doctor's professional fees, nursing care, social work services, and community paramedicine visits. Mount Sinai excluded noncore services, including inpatient consultations and post-acute radiology services, from the bundle and proposed that these services be billed separately as FFS. The bundled payment amount was set at 95 percent of the sum of the diagnosis-related group (DRG) payment[18] plus the estimated professional payments that would have occurred if the patient had been hospitalized.

[17] There is a partial payment structure for patients who only receive observation unit–level care at home or have to be escalated to a higher level of care during their HaH episode.

[18] DRG is a categorization system used to classify patients clinically and by level of expected resource use. DRGs are used by some insurance plans, such as Medicare Part A, to provide payment to hospitals for beneficiaries' inpatient hospital stays. More information can be found at these two websites (1) https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/index.html and (2) https://oig.hhs.gov/oei/reports/oei-09-00-00200.pdf.

The Medicare FFS model included a shared risk component. The HaH organization would be eligible for shared savings or responsible for repayment of a percentage of the annual difference between the total spending in HaH episodes and the benchmark. The benchmark is defined as the total expenditures of a comparison group (that is, a group of non-MACT patients who are similar to the MACT patients) minus a 3 percent discount (which reduces the comparison spending to a lower figure). The specific shared saving or repayment amount would be prorated based on the organization's performance on 10 quality metrics related to processes of care, beneficiary experience, safety, and functional outcomes. The total shared saving or repayment amount would be subject to a cap of 10 percent of the benchmark.

Mount Sinai developed additional payment models for other payers, which varied slightly to meet each payer's preferences. For example, Mount Sinai created a payment model for Healthfirst, an insurance company partially owned by Mount Sinai. The payment model designed for this payer covered the acute care phase of the HaH arm and the payment was structured as a physician global fee priced at a similar dollar amount as a discounted DRG. Unlike the model Mount Sinai designed for Medicare FFS, the Healthfirst model did not include the post-acute care phase nor did it include a shared risk component.

C. Status of the payment model

Mount Sinai has made progress developing, negotiating, and implementing its payment models. Mount Sinai executed a contract with Healthfirst for a bundled payment described above that covers core MACT services provided to patients during the acute care phase of the HaH arm. The payment model allows for MACT service provision to Healthfirst's Medicare Advantage and Medicaid managed care beneficiaries.

Mount Sinai formed a joint venture with Contessa Health, a company that partners with payers and providers to operate prospective bundled payment programs. Starting in October 2017, the joint venture began managing negotiations with payers and establishing payer-specific bundled payment models. Once these contracts are executed, the Mount Sinai-Contessa joint venture will receive payments from payers and process claims for the MACT program. These negotiated contracts may include the payer's Medicare Advantage, Medicaid managed care, and commercial plans. In fall 2017, MACT program leaders reported that the joint venture was in negotiations with approximately six payers with whom they hoped to reach agreements during 2018. While one or two initial payment model contracts may be executed before Contessa Health assumes the primary negotiation role, Mount Sinai plans to transition all MACT payment contracts to the Mount Sinai–Contessa joint venture in the long term. Mount Sinai leaders expressed their willingness to let other organizations join to help them overcome challenges in establishing payment arrangements with commercial payers.

D. Factors associated with the development of the payment model

Over the course of the cooperative agreement, Mount Sinai encountered and overcame barriers to establishing payment models and executing contracts with payers. Mount Sinai worked with payers that offer Medicare Advantage, Medicaid managed care, and commercial plans, to develop payment models. In-depth discussions with payers helped Mount Sinai identify challenges and questions regarding the establishment of HaH payment models. For example, payers had to consider whether they needed to add HaH as a new benefit through the Medicare

Advantage bidding process in order to cover HaH services for their Medicare Advantage beneficiaries. To overcome the challenges, Mount Sinai formed the joint venture with Contessa Health to leverage its experience in developing prospective bundled payment models for home hospitalizations.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Mount Sinai's main strategy to sustain the MACT program was to develop payment models so that payers had a mechanism to cover MACT services. The awardee implemented a payment contract with one payer (Healthfirst) and developed a joint venture with Contessa Health to establish payment processes and execute contracts with additional insurance companies. In addition, the Mount Sinai Health System committed to provide supplemental program funding until additional contracts with payers are executed.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Mount Sinai's sustainability strategy included finding funding and reducing the cost of the program. Mount Sinai leaders were working with multiple payers to develop payment models, and also considered reducing the number of program staff to lower the costs of sustaining the program.

C. Implementing the SSR plan: progress and changes

Sustainability. Mount Sinai's payment model was its main strategy for sustaining the program after the end of the cooperative agreement. In addition to the progress described in Chapter V, Mount Sinai began planning for a pilot HaH program at the James J. Peters Veterans Administration (VA) Medical Center. This pilot was scheduled to begin in early 2018 and Mount Sinai planned to enroll 25 patients. This pilot was funded by the VA.

While RaH payment models are being negotiated, Mount Sinai planned to continue the RaH arm by providing services to the homebound Medicare FFS population. Mount Sinai would sustain these activities by using a combination of house call provider codes and billing the therapy, nursing, and social work services as a certified home health episode. [19]

Scalability. Although one of the aims of Mount Sinai's scalability efforts during the cooperative agreement was to expand enrollment, Mount Sinai also used these activities to identify and explore program expansion options. For example, in the last quarter of the cooperative agreement, Mount Sinai implemented a pediatrics at home pilot program to gain an understanding of how the MACT program could provide services to benefit this patient population. At the end of the cooperative agreement, Mount Sinai was designing a new MACT arm that would provide postoperative care at home. The goal of this arm would be to provide patients, who routinely require postoperative hospitalization, hospital-level care in their home. Furthermore, at the end of the eleventh program quarter, Mount Sinai began recruiting from a third Mount Sinai hospital. While this may have been an attempt to expand enrollment, it also offered Mount Sinai both an additional opportunity to expand the MACT program to other hospitals and to set up the recruitment process at additional hospitals so that when additional contracts are executed with insurers Mount Sinai has the structure established to recruit patients from three hospitals.

Replicability. Mount Sinai had another grant, through the John A. Hartford Foundation, to support its replication efforts. Through this funding, Mount Sinai developed an implementation manual that outlines the MACT program processes. Mount Sinai planned to use this manual to assist other organizations in implementing a HaH model.

D. Factors associated with progress toward implementing the SSR plan

Mount Sinai worked with insurance companies for over two years to construct viable payment models. Although Mount Sinai leadership experienced several challenges, the leadership team was confident by the end of the cooperative agreement that its joint venture with Contessa would result in additional contracts being executed with insurance companies for the HaH and RaH arms of the MACT program.

[19] PT, occupational therapy, and speech therapy services would be billed with nursing and social work as a certified home health episode with FFS house calls.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. However, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or the sustainability, scalability, or replicability plans. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

Having concluded that we cannot estimate the impact of the HaH program, what analytic options are both available and useful? We have two responses. First, we believe it is worthwhile to understand the degree to which HaH participants resemble their counterparts who were admitted to inpatient care for the same condition. Although HaH participants are expected to be at lower risk than a typical inpatient, understanding the potential of the HaH program requires an attempt to quantify the differences between the two. To that end, we plan to compare Medicare-enrolled HaH participants and nonparticipants admitted to Mount Sinai with the same MS-DRGs in terms of prior spending and utilization, coexisting conditions, dual enrollment, and demographic characteristics.[20] Second, after further discussion with Mount Sinai project staff, we may use its control group to estimate differences in outcomes after discharge. As noted above, we remain uncertain that the two groups are sufficiently similar to characterize the differences as impacts. Moreover, the sample sizes of both groups are almost surely too small to detect any but the largest differences between groups. Nonetheless, results of these comparisons may generate useful information for follow-up research.

[20] It is unclear at this point whether all HaH participants were assigned a DRG. We are pursuing this question with project staff at Mount Sinai.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Mesa Fire and Medical Department

April 18, 2018

KeriAnn Wells (Mathematica Policy Research)
Margaret Gerteis (Mathematica Policy Research)
Lori Timmins (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The Mesa Fire and Medical Department received a six-month no-cost extension until February 2018. It will use the extension to continue a limited version of the program's 911 response component and work to finalize payment approaches to sustain the program.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

The City of Mesa used funding from HCIA R2 to develop the Community Care Response Initiative (CCRI), designed to resolve several problems associated with the use of the 911 emergency response system for non-emergencies, including: (1) the high cost of transporting people with low-acuity conditions by ambulance and treating them in hospital emergency departments (EDs); (2) the diversion of professionals and resources needed to respond to true emergencies; and (3) EDs overcrowding with patients who do not need emergency care.

The CCRI addressed these problems by deploying community medicine (CM) units to provide non-emergency services as appropriate to 911 callers at their homes or in the community. The awardee introduced two types of CM units—CM medical units and CM behavioral units—to respond to low-acuity 911 calls, choosing the appropriate unit based on the callers' needs. CM medical units were ambulances staffed with a captain paramedic and an advanced practice provider (APP), who provided on-site services similar to those in an urgent care setting, such as suturing wounds and administering antibiotics. They also helped participants arrange follow-up appointments with their usual providers.

CM behavioral units were sport utility vehicles staffed with a captain paramedic and a licensed behavioral health clinician, who provided non-emergency behavioral health or crisis intervention services for conditions including anxiety, depression, substance abuse, and suicidal ideation. The APPs who worked on the CM medical units were officially employed by Mountain Vista Medical Center in Mesa, but worked alongside fire department paramedics and reported to battalion chiefs in the fire stations for operational issues. Behavioral health clinicians on the CM behavioral units were employees of Crisis Preparation and Recovery, Inc., (CPR), but they, too, worked out of the fire department and reported to the battalion chiefs for operational issues.

To complement the 911 response component of the CCRI, which focused on low-acuity cases, the Mesa Fire and Medical Department added a care transitions component that provided home-based services to higher-acuity participants with targeted conditions following a discharge from the hospital. The care transitions component also paired an APP and a captain paramedic, who traveled to participants' homes in a CM medical unit within 72 hours of discharge. Care transitions teams provided post-discharge services such as medication reconciliation and referral to other clinicians. The care transitions component was secondary to the 911 response component, to which the majority of HCIA R2 funding was allocated.

Figure I.1. The Mesa Fire and Medical Department: Community Care Response Initiative program components

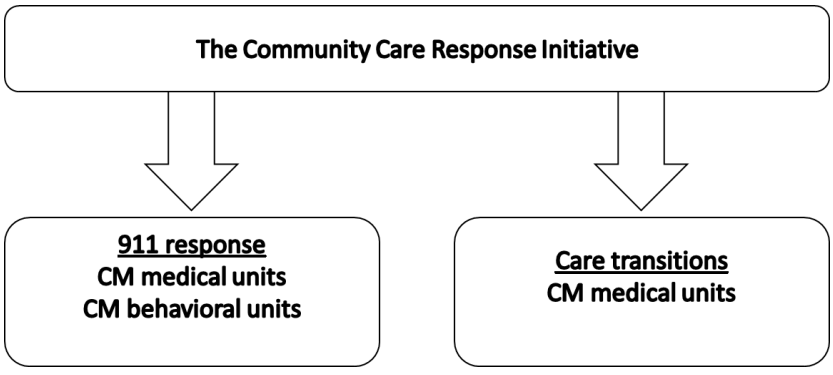


Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The Mesa Fire and Medical Department redesigned its 911 emergency response by dispatching CM units to low-acuity callers and treating them on-site in the community—often at home—instead of transporting them to the ED. CM units also provided care transition services to participants with selected chronic conditions who were recently discharged from the hospital.
Major innovation	Adding non-emergency clinical response teams into 911 dispatch protocols
Program components	• **911 response**. Using (1) CM medical units staffed with an APP and a paramedic or (2) CM behavioral units staffed with a licensed behavioral health clinician and a paramedic to provide direct care to 911 callers whose conditions did not warrant ED visits • **Care transitions**. Conducting home visits and care coordination for high-acuity participants within 72 hours of hospital discharge
Target population	• **911 response.** Low-acuity 911 callers • **Care transitions.** Patients with congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), myocardial infarction (MI), sepsis, and pneumonia who were recently discharged from Dignity Health hospital and identified as being at high risk for readmission
Theory of change/ theory of action	Mesa hypothesized that (1) using CM units to treat low-acuity participants on-site in the community would decrease both inappropriate ED use and inappropriate use of ambulances to transport participants to the ED, and (2) using CM medical units to provide higher-acuity participants with care transition services in their homes after hospital discharge would result in fewer readmissions to the hospital. This approach would help reduce ED overcrowding, focus emergency services on priority patients, and reduce hospital readmissions, thus reducing costs and improving the quality of care.
Payment model	New fee-for-service (FFS) payment
Award amount	$12,779,725
Launch date	December 1, 2014
Program setting	**911 response**. CM medical and behavioral units dispatched from Mesa Fire and Medical Department stations to community settings and participants' homes **Care transitions**. CM medical units dispatched from Dignity Health to participants' homes
Market area	Urban, suburban
Market location	Mesa, AZ
Target outcomes	• Reduce low-acuity patients' ED visits and ambulance use by 40 percent in three years • Reduce high-acuity patients' hospital readmissions in three years (high-acuity patients include those diagnosed with CHF, COPD, MI, pneumonia, and sepsis) • Reduce ambulance use and ED visits to save $41 million in three years

APP = advanced practice provider; CM = community medicine; ED = emergency department.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that Mesa Fire and Medical Department was partly successful at implementing its program by the end of the initial three-year cooperative agreement. Specifically, the awardee successfully implemented the primary component of 911 response, but was less successful implementing the secondary component of care transitions. We based this conclusion on five factors.

First, the awardee enrolled 12,431 participants—85 percent of its enrollment target—by the end of the initial cooperative agreement. About 98 percent of participants were enrolled in the 911 response component.

Second, the awardee successfully implemented the 911 response component early in the cooperative agreement, but made little progress implementing the care transitions component before the third year of the cooperative agreement.

Third, the awardee nearly met its staff recruitment goals and maintained high rates of staff retention until the third year of the cooperative agreement, when some APPs left because they were not sure they would have a job after the end of the award.

Fourth, the awardee successfully engaged participants in both the 911 response and care transitions components, earning positive comments on participant surveys and low refusal rates among participants using either the 911 response or the care transitions components.

Finally, participating clinicians and other implementation staff reported that they believed the program had a positive effect on care delivery, because it provided care in the comfort of patients' homes or transported them to appropriate behavioral health facilities, preventing them from waiting in EDs for non-emergent services.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Mesa Fire and Medical Department's CCRI program, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. The awardee used existing pre-hospital and behavioral health fee-for-service (FFS) billing codes to recoup a subset of the costs of providing community medicine services to 911 callers who do not need emergent care. The awardee engaged local payers throughout the cooperative agreement, and although payers saw the value of the program, they ultimately would not enter into shared savings or other types of agreements with the awardee, citing competing priorities and lack of a direct financial incentive. The awardee has approached CMS about adopting pre-hospital billing codes that cover the costs of paramedics on CM behavioral units and APPs on CM medical units.

Sustainability. The awardee will discontinue the care transitions component and maintain a scaled-back version of the 911 response component during the no-cost extension. The awardee hopes to sustain the 911 response component after the extension—however, that will depend upon whether the Mesa Fire and Medical Department can be adequately reimbursed. Other fire and medical departments, such as those in the cities of Anaheim and Los Angeles, California, have used local funds to implement similar community medicine programs for 911 callers with low-acuity issues.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July to October of 2016 with a sample of 20 potential respondents and achieved a response rate of 95 percent. The clinician survey was fielded from March to June of 2017 with a sample of 11 potential respondents and achieved a response rate of 91 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that Mesa Fire and Medical Department was partly successful at implementing its program. Specifically, the awardee successfully implemented the 911 response component, but was less successful implementing the care transitions component. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

911 response component. For the 911 response component of its intervention, the awardee identified patients eligible for low-acuity CM services by using dispatch protocols designed to rule out true or possible emergencies. As experience taught the awardee which calls were most appropriate for low-acuity CM services and which were not, the protocols were revised accordingly. Participants were considered enrolled in the 911 response component if a CM unit was dispatched to their location after a 911 call, even if the participant ultimately took an ambulance to the ED. In addition to fielding 911 calls from community members, local police also requested CM units to help them work with citizens with behavioral health issues and jailed

citizens who needed medical clearance. Emergency responders also requested CM units when they encountered citizens in the field who did not require emergent care.

Care transitions component. In December 2016, the end of the second year of the HCIA R2 collaborative agreement, the awardee relocated the care transitions component from Mountain Vista Medical Center to a Dignity Health hospital and changed the eligibility criteria. Patients at Mountain Vista Medical Center and (later) at Dignity Health were considered eligible for the care transitions component if they were being discharged with at least one identified high-risk condition. At Mountain Vista Medical Center, the awardee initially focused on enrolling patients with congestive heart failure. After transitioning to Dignity Health, in an effort to boost enrollment, the awardee expanded the list of conditions to include acute myocardial infarction, chronic obstructive pulmonary disease, sepsis, and pneumonia.

> "I think people really trust [emergency medical services (EMS)] and EMS workers, and I think most people will always let them in their home the majority of the time, just because they feel safe."
>
> —Dignity Health employee

At both facilities, hospital staff initially identified and enrolled participants, but when enrollment remained low, the Mesa team assumed that responsibility. At Dignity Health, hospital staff gave the care transitions team a list of patients who lived in the awardee's response region and had at least one of the targeted conditions. The awardee team visited patients before they were discharged, introduced the program to them, and asked them to participate. Both hospital and awardee staff reported that patients were more likely to enroll when they were approached by fire department personnel and less likely to be responsive to hospital personnel, and attributed this difference to the public's general tendency to trust first responders. The care transitions team said that most patients they approached agreed to enroll in the program, particularly after they found out it was free.

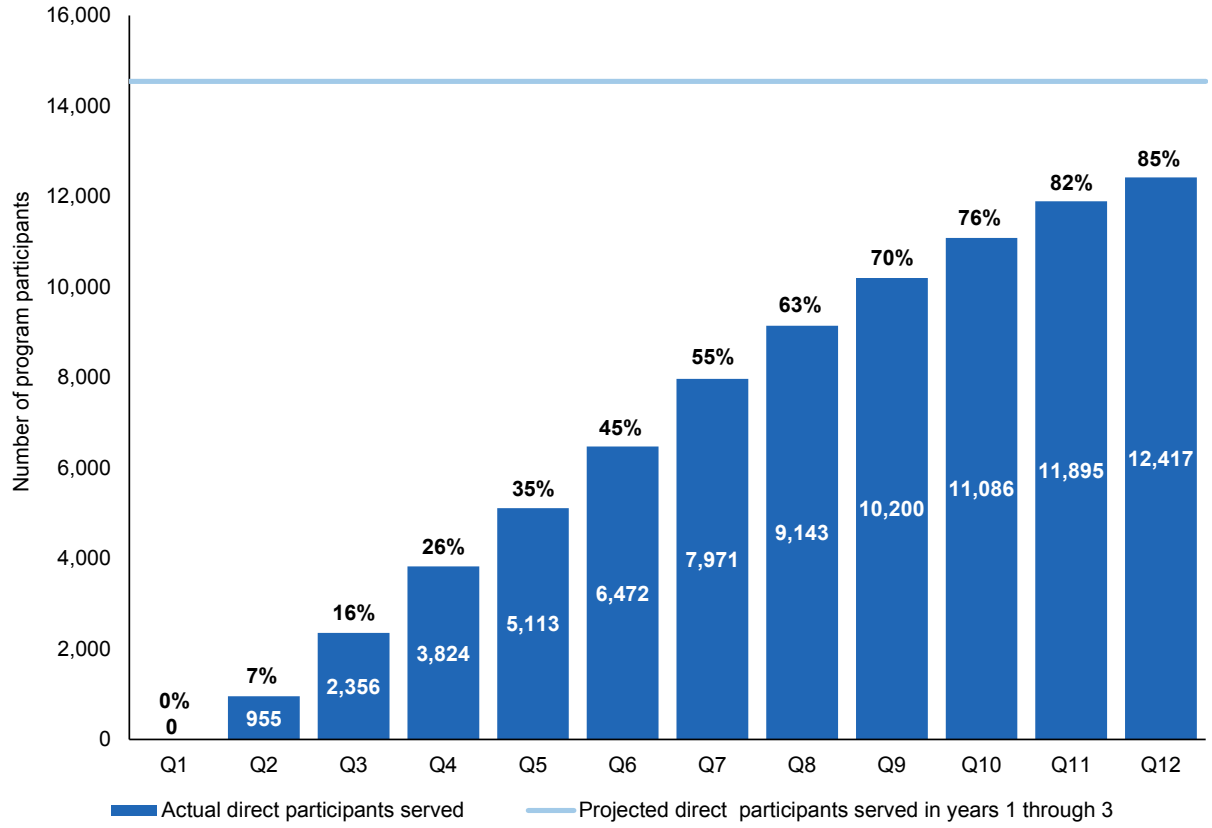
b. Evidence of enrollment effectiveness

Overall, the awardee reported that it enrolled 12,431 participants from December 2014 (when it launched its program) through August 2017, which represents about 85 percent of its final three-year projection (Figure II.1). The awardee originally aimed to serve 27,179 patients, but revised its target in spring 2016 to 16,200 patients based on actual participants served. In fall 2016, the awardee further revised its target enrollment to 14,543 patients. Projections are for both the 911 response and care transitions components; the awardee did not project component-specific enrollment. Enrollment decreased in the final half of the cooperative agreement, from a high of 532 participants served in March 2016 to a low of 141 participants served in August 2017. About 98 percent of participants were enrolled in the 911 response component. The awardee reported serving a total of 260 participants in the care transitions component, the majority of whom were served between September 2016 and June 2017. In survey results, 94 percent of non-clinician staff and more than half of clinicians agreed or strongly agreed with the statement: "To date, we have successfully recruited or enrolled the targeted number of participants into the program."[4] Given these results, we conclude that the awardee was partially

[4] The staff survey was conducted in the fall of 2016, and the clinician survey was conducted in the spring of 2017. Therefore, staff responses are based on their perspectives two years into the three-year cooperative agreement, whereas clinician responses are based on their perspectives at the end of the cooperative agreement.

successful recruiting patients to its program, and had much more success recruiting patients to the 911 response component than to the care transitions component.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee originally aimed to serve 27,179 patients. In spring 2016, it revised its target to 16,200 patients based on actual participants served. In fall 2016, the awardee further revised its target enrollment to 14,543 patients.

c. Barriers and facilitators associated with enrollment effectiveness

Mesa Fire and Medical Department's progress in meeting its three-year enrollment goals was influenced by two key factors: (1) community engagement in the 911 response component and (2) the limited capacity of the care transitions team.

911 response component. The awardee successfully engaged both internal and external stakeholders to raise awareness of the 911 response component's value to the community over the course of the HCIA R2 cooperative agreement. Initially, emergency response staff at non-participating fire stations rarely requested CM units when they encountered low-acuity patients in the field. In response, the awardee continuously reached out to Mesa Fire and Medical

Department fire crews at stations lacking CM units, and moved some captain paramedics to these stations following their tenure on CM units. These efforts increased awareness about the program and improved buy-in among the organization's personnel. The awardee also educated local police and schools about the program. Awardee staff said these activities resulted in more referrals to the program from Mesa emergency response teams at non-participating stations, as well as from police and schools.

The awardee welcomed referrals from police, but about halfway through the cooperative agreement, awardee leaders expressed concerns that police were requesting CM units too often. In response, the awardee worked to educate police about which patients were most likely to benefit from CM units. Furthermore, in the final year of the cooperative agreement, the police department implemented a new program of its own that improved its ability to manage and transport citizens with behavioral health issues. These two factors might have reduced the number of police requests for CM units, but the appropriateness of the police requests that were received reportedly improved.

Care transitions component. The awardee's decision to move the care transitions component to Dignity Health from Mountain Vista Medical Center took place after a change in leadership at Mountain Vista Medical Center, a persistent lack of buy-in among hospital staff, and low enrollment levels. Dignity Health is a larger hospital network than Mountain Vista Medical Center, with an existing care transitions program and more patients in the awardee's response region than Mountain Vista Medical Center had. As noted, the care transitions team assumed the role of identifying and enrolling patients in the hospital. Although the care transitions team believed that directly recruiting patients increased enrollment in the program, time spent recruiting and enrolling patients in the hospital left less time for the team to conduct home visits. The team believed that having a dedicated scheduler might have enabled it to make more home visits.

2. Delivery of program services

a. Description of and changes to service delivery model

911 response component. The awardee's service delivery model entailed dispatching CM units to address low-acuity needs of 911 callers in an effort to divert them from the ED. "Low-acuity" medical issues commonly included self-limiting illnesses (such as flu), minor injuries requiring sutures, diabetes management, management of nose bleeds, or conditions requiring antibiotics. APPs on CM medical units offered on-site evaluation, treatment, and referrals to primary care physicians or specialists. CM behavioral units responded to participants with issues such as agitation, depression, or suicidal ideation who would be unlikely to benefit from an ED visit. CM behavioral unit visits often lasted a few hours as behavioral health clinicians assessed participants and worked to find the most appropriate treatment venue. Many behavioral health calls resulted in patients being directly transported to local inpatient behavioral health facilities, bypassing the ED.

To support its 911 response component, the Mesa Fire and Medical Department redesigned and continually refined its 911 dispatch protocols to incorporate the new low-acuity response options using CM units. The awardee's dispatch center served all stations in the region, including the three stations in Mesa and one station in Apache Junction, Arizona, that were participating in

HCIA R2. The Mesa Fire and Medical Department initially hired two triage nurses to provide backup support to dispatchers in deciding when to deploy CM units. Triage nurses were available at the dispatch center to give telephone support to patients while CM units were en route. In practice, however, the awardee found that triage nurses were underutilized, because it was usually necessary to resolve doubt about the acuity level of a call on-site. In the spring of 2017, the awardee discontinued the nurse triage function due to low demand for its services.

When the awardee launched the program, it deployed three CM medical units and three CM behavioral units. One CM medical unit operated out of neighboring Superstition Fire and Medical Department in Apache Junction, Arizona. The other units operated out of three Mesa Fire and Medical Department stations. In the final year of the award, in response to lower demand for services in Apache Junction, the awardee relocated the CM medical unit from Apache Junction to Mesa, effectively ending the 911 response component at Superstition Fire and Medical Department.

In April 2017, in response to attrition among APPs and in an attempt to concentrate services during peak hours, the awardee changed CM medical units' hours of operation. All CM medical units transitioned from providing services 24 hours a day, 7 days a week, to operating from 8 a.m. to 10 p.m., 7 days a week. This enabled the awardee to offer APPs and captain paramedics three 13-hour shifts a week instead of rotating 24-hour shifts, and APPs credited this switch with increasing their job satisfaction. The awardee also shifted one CM behavioral unit to operate 9 a.m. to 7 p.m. Monday through Friday; the other two CM behavioral units continued operating from noon to 10 p.m., 7 days a week. Behavioral health clinicians reported that covering these weekday morning hours enabled the awardee to better meet the demand for services. As of July 2017, three CM medical units and three CM behavioral units were operating in Mesa. Awardee reports through August 2017 suggest that during the no-cost extension period, a total of two CM medical units and one CM behavioral unit were operational.

Care transitions. Service delivery for the care transitions component consisted of the care transitions team visiting patients at home within 72 hours of their discharge from the hospital. The care transitions team consisted of an APP and captain paramedic who were available Monday through Thursday, 7 a.m. to 5 p.m. Typical services included medication reconciliation, care coordination with home health or primary care providers, and home safety assessments. Because the team worked Monday through Thursday, it sometimes waived the 72-hour requirement—for example, when patients were discharged on a Friday. In December 2016, when the awardee relocated the care transitions component from Mountain Vista Medical Center to Dignity Health, it also reduced the number of teams dedicated to the component from two to one.

b. Evidence of service delivery effectiveness

The awardee effectively implemented the 911 response component, continuously improving the quality of the dispatch and response process throughout the cooperative agreement. Staff believed in the program's value to participants and to the health system as a whole, and participants also expressed appreciation for the program. The awardee was partly successful in implementing the care transitions component, achieving the most success in the final year of the award.

Delivery of intervention services. In the first quarter of the cooperative agreement, the awardee developed operational plans, secured CM vehicles, and hired APPs and behavioral health clinicians. The awardee launched the 911 response component in the second quarter of the cooperative agreement, effectively dispatching both CM medical units and CM behavioral units to treat participants in the community with low-acuity conditions.

Throughout the cooperative agreement, the awardee reviewed internal monitoring and tracking data on the 911 response component to continuously refine its dispatch protocols and treatment protocols and ensure that CM teams responded to truly low-acuity situations in which patients were most likely to benefit from on-site services and avoid an ED visit. For example, dispatchers stopped sending CM medical units to patients with abdominal pain because they could not divert most of those patients from the ED. In the final year of the award, the awardee also changed its dispatch protocol so that CM behavioral units only responded to calls directly from patients or from callers physically in the same place as patients. Before the change, CM behavioral units could respond if a concerned friend or family member who was not in the same place as the patient called 911, but this made it more difficult to accurately determine the patient's acuity.

The implementation of the care transitions component was partially successful. The awardee intended to launch the component at Mountain Vista Medical Center in 2015, the first year of the cooperative agreement, but only reported serving one participant that year. Enrollment in the care transitions component increased in the final quarter of the second year, and remained steady after the component was transitioned to Dignity Health.

Transitioning the program to Dignity Health in the final year of the award had advantages and disadvantages. The APP and captain paramedic applied knowledge gained at Mountain Vista Medical Center to their approach at Dignity Health, taking responsibility for identifying and enrolling patients and scheduling home visits. However, the APP on the care transitions team, who was an employee of Mountain Vista Medical Center, found it difficult to coordinate care planning with hospital staff at Dignity Health because he was not officially employed there. For example, he could not attend staff meetings at Dignity Health as he did at Mountain Vista Medical Center. In lieu of staff meetings, the awardee team communicated informally with Dignity Health care coordinators and other staff to discuss potentially eligible patients who might benefit from the Mesa team's care transition services.

Staffing and training. The awardee successfully recruited 11 APPs and four behavioral health clinicians when it launched the 911 response component, achieving 88 percent of its staffing target by the end of the first year of the cooperative agreement. The awardee initially planned to hire two more triage nurses and two additional CM medical unit teams. However, the awardee did not hire the triage nurses because nurses were receiving fewer calls than expected, and most calls resulted in dispatching either a CM unit or emergency response. In the second year of the cooperative agreement, Mesa Fire and Medical Department's partners hired an APP and a behavioral health clinician for the program, but decided not to hire another CM medical team as originally planned.

Staff retention was strong until the final year of the cooperative agreement, when some APPs left because they did not know whether they would continue to be employed after the end

of the cooperative agreement. As APPs left the program, the awardee reduced the number of CM medical teams, which meant that captain paramedics who had been paired with APPs returned to traditional emergency response services. The awardee also reduced the hours of CM medical unit operation, as noted. CPR assured behavioral health clinicians they would continue to be employed regardless of whether the awardee sustained the program after the cooperative agreement, resulting in less attrition among behavioral health clinicians.

> "We've had pretty much the same staff, I think, through most of the grant, but especially this last year. I think it just attributes to how much we love this job and love what we're doing and how we can bring change every single day, [which] is what we all love most about this, and I think that's what helps to retain us."
>
> —Behavioral health clinician

During our discussions with the awardee in June 2017, the awardee had three behavioral health clinicians working on CM behavioral units, four APPs working on CM medical units for the 911 response component, and one APP working on the care transitions component. After our discussion, awardee reports from the final quarter suggest that the awardee retained one behavioral health clinician and two APPs, thus paring down the program for the no-cost extension period.

To train staff, the awardee offered 24 hours of crisis intervention training to all staff. Throughout the award, the awardee also offered short, specialized training sessions on topics like suture and wound care, epistaxis care, and coordination with jails to medically clear inmates. Captain paramedics new to the program shadowed other captain paramedics on CM medical and behavioral units for two to three weeks before they started working independently with a clinician.

In the third year of the cooperative agreement, the awardee offered a two-day APP skills training workshop to achieve more uniformity of skills across the APPs in providing different types of urgent care, such as catheter replacement. In the workshop, APPs developed their competence on all core services. Clinicians and other staff also reported various types of informal training achieved through attending staff meetings, asking colleagues for help, self-study, shadowing other staff, and getting technical assistance.

By the end of the three-year cooperative agreement, the awardee had reported offering a total of 943 hours of training.

Recruitment and engagement of providers. The awardee's staffing for the CCRI included APPs (nurse practitioners and physician assistants) employed by Mountain Vista Medical Center, who are included in the discussion of staffing and training above. However, the intervention otherwise did not involve much provider recruitment or engagement.

Engagement of program participants. Mesa staff reported that the majority of patients were very satisfied with the program, especially the 911 response component, which saved them an ambulance trip to the ED. In surveys, 94 percent of staff and more than three-quarters of clinicians agreed or strongly agreed with the statement: "To date, we have successfully engaged participants with the program."

> "We were so pleased to have the medical staff come to our home rather than being transported to another facility. The care and treatment [were] professional, efficient, and understanding of our concerns. We are grateful for their attentiveness and thankful that they are available to us."
>
> —Patient survey respondent

The awardee also sent surveys, adapted from the Hospital Consumer Assessment of Healthcare Providers and Systems, to patients after their encounters. There were 301 responses through August 2017. Although the survey was based on a convenience sample and therefore limited by selection bias, responses to it were overwhelmingly positive. For example, 84 percent of patients rated the CM service as "excellent," and 10 percent said it was "very good." Almost all (97 percent) said they would recommend the service to friends and family.

> "You will get some percent[age] of patients, possibly with an addiction to narcotics, that definitely want to go to the hospital. Those are the ones that usually do not want to participate if they already have it in their mind that they want a specific medication at the hospital. Sometimes you see it with the homeless when [the weather is] hot."
>
> —APP

Some staff expressed frustration about participants who refused 911 CM services and insisted on taking an ambulance to the ED.[5] The awardee's self-reported measures suggest that about 3 percent of patients refused services between June 2015 and August 2017. Staff and clinicians, responding to the survey and interviews, suggested that patients who refused CM services often either wanted to go the ED to receive certain medications, especially narcotics for pain, or were homeless and seeking food and shelter in the ED.

c. Barriers and facilitators associated with service delivery effectiveness

Mesa Fire and Medical Department's effectiveness at service delivery was influenced by five factors, including: (1) program buy-in among staff, clinicians, and the community; (2) local policy changes; (3) communication and coordination with clinicians and partners; (4) integrating clinicians into fire department culture; (5) documentation challenges.

First, throughout the cooperative agreement, staff and clinicians repeatedly expressed their support for the 911 response component as a better model for providing care. During interviews, staff characterized this buy-in among paramedics and clinicians as a major factor in the program's success, allowing them to promote the program effectively to participants and community partners, such as police and other Mesa Fire and Medical Department staff. In surveys, clinicians and staff cited buy-in as one of the

> "It's very fulfilling to [tell a patient], 'Well, we can take you right to a behavioral hospital now and see a psychiatrist right away.' That's a huge thing to be able to see people … access the mental health care that they need in the moment."
>
> —Behavioral health clinician

[5] Patients who refused services were still considered enrolled in the program because a CM unit was dispatched to their location. They were not considered successful ED diversions.

most helpful factors to CM units achieving their goals. The awardee never accomplished the same level of buy-in and engagement with the care transitions component, however. In interviews, some staff suggested there was more buy-in to the 911 response component because that component was integrated directly into Mesa Fire and Medical Department's response protocols. In contrast, the care transitions component was located in the hospital and largely isolated from the fire department and emergency response system.

The second factor that improved service delivery effectiveness occurred early in 2017, when the city of Mesa authorized the awardee to transport non-emergent patients to the ED in CM medical units for additional testing, such as x-rays. Previously, the awardee was required to dispatch a third-party ambulance that would transport patients and bill payers for that service. This change improved care coordination because APPs could communicate directly with clinicians at the hospital instead of communicating through ambulance operators. It also enabled the awardee to receive compensation for transportation.

Third, communication and coordination with APPs and partners was a persistent barrier to implementation, and problems intensified somewhat in the final year as APPs began to worry about keeping their jobs after the end of the cooperative agreement. In interviews conducted throughout the cooperative agreement, multiple staff cited communication challenges with partnering organizations and between the awardee and clinicians as a barrier to program implementation. When asked in surveys about barriers encountered by the CM medical and behavioral units in achieving their goals, the majority of clinicians and staff selected "difficulty working across collaborating organizations," "ineffective communication," and "challenges integrating program into existing structures" as either major or minor barriers. Furthermore, a majority of clinicians who made general comments in the survey cited challenges communicating with clinicians and coordinating across entities. Their comments suggested that they did not always consider Mesa Fire and Medical Department as open to input from clinicians. Collectively, these findings suggest that Mesa Fire and Medical Department leaders could have more effectively engaged clinicians and partners. However, behavioral health clinicians did not cite communication with the awardee as a challenge, perhaps because CPR was reportedly more engaged with its employees and willing to offer them post-award job security.

Fourth, Mesa Fire and Medical Department staff also faced challenges integrating clinicians into their culture, which is protocol-oriented and has a clear hierarchical chain of command. Some chiefs expressed concerns about having clinicians in their station who did not report to them, and captain paramedics noted that all APPs did not have the same skill set in the field. In response, the awardee began having clinicians report to battalion chiefs for certain operational issues in the second year of the award. In addition, in the third year of the award, the awardee offered APPs a two-day training to develop uniform competency on core services.

Fifth, the awardee faced challenges related to documentation. Initially, the awardee documented 911 response encounters in an encounter-based, emergency medical services (EMS) tracking system, which was not conducive to participant tracking and billing. After a lengthy procurement and legal review process with the city of Mesa, in May 2016 the awardee implemented an electronic medical record (EMR) for APPs, while captain paramedics continued

to document encounters in the EMS system.[6] However, the awardee continued to face challenges with the new EMR and found it was not conducive to pre-hospital care.

In December 2016, Mesa Fire and Medical Department shifted back to its original system, with some enhancements to accommodate billing. This shift reduced documentation workloads, enabling APPs and paramedics to document in the same system. These documentation challenges might also have contributed to persistent quality issues in the data the awardee submitted for the impact analysis. Participants' insurance status was often missing or incorrect, for example, limiting our ability to identify participants in claims and to analyze the program's impacts on participants with low-acuity conditions.

C. Assessment of perceived program effects on the delivery of care and outcomes

Between March 2015 and August 2017, the awardee reported diverting 7,359 people who would otherwise have taken an ambulance to the emergency room. This represented about 65 percent of participants in the 911 response component. Program leaders and staff noted that behavioral health units were especially successful, estimating that the diversion rate was about 80 percent for participants to whom CM behavioral units were dispatched. In responding to surveys, more than three-quarters of clinicians and non-clinician staff thought the CM units had been somewhat or very effective in achieving their goals.

Leaders and staff were less certain about the care transition component's success in reducing readmissions. A Dignity Health staff member said the readmission rate among participants was about the same as it was for the hospital as a whole, but noted that the hospital planned to do more research to compare outcomes among patients with high-acuity conditions. Similarly, Mesa leaders said they had not observed any reduction in readmissions from the care transitions component. However, both parties believed that a shorter-term 15-day readmission rate would be a more relevant metric than the 30-day readmission rate because teams only saw the patient once within 72 hours of discharge, and their goal (according to awardee leaders) was to bridge the gap between discharge and home health, which starts two weeks after discharge. Staff believed that any success in reducing readmissions was primarily due to the care transitions teams' medication reconciliation services. The care transitions team also discussed cases in which, during the post-discharge home visit, they recommended that patients return to the hospital, and they believed this prevented the patients from being readmitted later with more severe symptoms.

In interviews and surveys, respondents also identified several intermediate effects of the CCRI on care, such as improving the quality of care. In the staff survey, 93 percent said their ability to provide care or services to participants was better than it was before the cooperative agreement, and 89 percent said the program was making a difference in meeting critical needs in the community. On every aspect of care they were asked to comment on in the survey, the

[6] Throughout the cooperative agreement, behavioral health clinicians used CPR's documentation system, because their licenses require extensive documentation of patient encounters.

majority of staff reported positive impacts (Table II.2). Nearly all clinicians also perceived positive impacts on the aspects of care listed in Table II.2.[7]

Table II.2. Percentage of clinicians and staff responding that CCRI had a positive impact on selected care goals

Please indicate the impact you believe the CCRI has had on each of the following:	Percentage of staff selecting "positive"
The quality of care and services you provide to participants	100
Your ability to respond in a timely way to participant needs	100
Your ability to provide care or services that are responsive to participant preferences, needs, and values	88
Access to care or services for all participants	81
Achievement of participant health goals	63
Participant satisfaction	100
Participant quality of life	94
Care coordination	100
The efficiency of care or services provided to participants	100
Care or services that are provided fairly to all participants	89

Source: Staff surveys.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

Based on findings from our implementation evaluation, and assuming the awardee's theory of action is correct, the awardee expects to see less ED use at Mesa facilities during the cooperative agreement relative to before the award. Given lower enrollment in the 911 response component at the Apache Junction site, it might be more difficult to detect an effect in ED visits at facilities in and near Apache Junction. Furthermore, the awardee theorized that 911 response teams would refer some participants to more appropriate services, such as primary care or specialty services. However, this might not have happened as often as intended, especially among participants who were served outside of business hours, lacked a usual source of care, or were visiting retirees whose usual providers were out of state.

The awardee's goal was to reduce ED visits by 40 percent among low-acuity participants, not the system as a whole. In its original application, the awardee estimated that nearly 43 percent of ambulance transports were for low-acuity conditions, which translates to a goal of reducing overall ED visits by 17 percent compared to baseline.

It might be difficult to detect program effects given persistent data quality problems that prevent us from identifying many program participants. Any observed reductions in ambulance

[7] Two options offered in the staff survey shown in Table II.2 were not options on the clinician survey: "the efficiency of care or services provided to participants" and "care or services that are provided fairly to all participants."

use and ED visits might also result in cost savings, although awardee leaders commented that savings will be far less than their $41 million goal. They estimated about $13 million in savings through February 2017. The awardee also intended to reduce hospital readmissions by a significant margin, but the care transitions component did not serve many participants and it might not be possible to observe whether this reduction took place. However, any readmissions reductions would probably be concentrated among patients with targeted conditions who were discharged from Dignity Health in the final year of the cooperative agreement.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Mesa Fire and Medical Department's CCRI program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Mesa Fire and Medical Department received a six-month no-cost extension and will continue to operate until February 28, 2018. The projected sample sizes in Table III.1 are based upon actual enrollment as of August 31, 2017, projected through the extension period, and will likely be the maximum number of beneficiaries that will be included in our evaluation. Due to processing lags in Medicaid data, we have not confirmed that all 3,000 Medicaid beneficiaries meet program eligibility for inclusion in our impact evaluation.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: Mesa Fire and Medical Department

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population at program end	803[a]
Projected Medicaid population at program end	3,000[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,671
Likelihood of all-cause hospitalizations	972
MDE sample size requirement to detect 20% effect	
Total expenditures	418
Likelihood of all-cause hospitalizations	243
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Questionable, no testing yet to determine strength of intent-to-treat framework
Likelihood of solid comparison group	Some issues, but probably surmountable; expect to select a comparison group

Table III.1 *(continued)*

Evaluability domain	Response
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff, clinician, and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We anticipate conducting a rigorous impact evaluation of the CCRI program by using difference-in-differences estimation with propensity score matched comparison groups for Medicare and Medicaid beneficiaries. The CCRI targets two groups of beneficiaries for two different components of the intervention. The first is low-acuity 911 callers with targeted conditions, including acute illnesses such as colds and flus, mental disorders or behavioral health issues, substance abuse disorders, and health-related behaviors such as stress management and medication adherence. The second component of the initiative targets high-acuity patients with CHF, COPD, MI, sepsis, and pneumonia who were recently discharged from Dignity Health. Although both program components deploy mobile CM units to participants' homes, the services differ based on needs. Based on current enrollment trends, we expect to have a sufficient sample size to analyze the low-acuity, 911-caller component but not the high-acuity care transitions component. Consequently, we will focus on the 911-caller component in the impact evaluation. We will choose comparison group communities that have similar demographics, medical care services, and socioeconomic characteristics as Mesa, Arizona, where the awardee operates. We will then use propensity score matching of patients from these communities to the treatment beneficiaries in order to derive that comparison group that is similar in terms of medical, payer, and demographic information. We believe we have a sufficient number of Medicare and Medicaid participants to detect a 20 percent effect or smaller on Medicare and Medicaid expenditures.

B. Characteristics of Medicare and Medicaid participants at baseline

The Mesa Fire and Medical Department's CCRI treatment group consists of the 911 callers with low-acuity issues who received care from a CM unit rather than from an emergency response team that would otherwise transport the caller to an ED by ambulance. The decision to send the CM unit rather than an ambulance is made by the 911 operator, with the help of a computerized decision tree. The CM unit staff evaluate the caller; treat him or her if necessary; and direct the caller to further, appropriate care—also as necessary. For callers with behavioral health problems, a behavioral health counselor in the CM behavioral unit evaluates and directs callers to appropriate care. The awardee considers an individual encounter to be a success when a low-acuity patient is not transported by ambulance to the ED.

For the purposes of presenting baseline characteristics in this report, we restricted the treatment group to Medicare beneficiaries who were seen by a CM medical or behavioral unit

before May 31, 2016.[8] We used the individual's first encounter with the CM unit in this period to trigger each person's baseline year. These beneficiaries had to be enrolled in Medicare FFS, both Parts A and B, for at least 90 days in the baseline year (the 365 days immediately before their enrollment). We also had to be able to link them to Medicare claims and enrollment data. The calendar period covered by the baseline quarters is based on the enrollment date for each enrollee and therefore varies by enrollee. After we excluded beneficiaries who did not meet the above criteria, a total of 629 participants were included in the analysis of baseline characteristics for this report.[9]

Table III.2 shows the demographic and health status characteristics of 911 callers who are Medicare FFS beneficiaries. A sizeable proportion of this group is younger than 65 (29 percent), which indicates that the treatment group of Medicare FFS beneficiaries is relatively young compared with the broader Medicare population. Most participants are white (89 percent), with slightly more females (55 percent) than males (45 percent) in the sample. The majority of participants originally qualified for Medicare because of their age (59 percent). Nearly one-third (30 percent) of the beneficiaries are dually eligible for Medicare and Medicaid, compared with 18 percent in the general Medicare FFS population—although, in some cases Medicaid benefits may be restricted to the payment of co-insurance and deductibles. The average hierarchical condition category (HCC) risk score of participants is 2.24—more than double the average for Medicare FFS beneficiaries nationwide (1.0). Although participants with very high HCC scores are driving up the average in the sample, more than half of these participants have HCC risk scores that are higher than the national average. In addition, although the treatment population grows with each quarter, the baseline demographic characteristics are remarkably similar across the three quarters that we have analyzed—which suggests that the Mesa Fire and Medical Department's intervention benefits a specific type of Medicare FFS beneficiary. Participants in the Mesa Fire and Medical Department's CCRI have substantially poorer health status and a greater need for care than most Medicare FFS beneficiaries.

[8] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

[9] A beneficiary may have had multiple encounters with the mobile health units after the program was launched. The awardee regards the encounters as separate cases in its participation counts. However, for the purpose of producing baseline characteristics, we considered only unique beneficiaries seen by a mobile health unit in the post-period.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Mesa Fire and Medical Department's program through May 31, 2016

	All participants (N = 629)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	184	29
65 to 74	180	29
75 to 84	145	23
85 and older	120	19
Gender		
Female	344	55
Male	285	45
Race		
White	559	89
Black	27	4
American Indian, Alaska Native, Asian/Pacific Island American, or other	23	4
Hispanic	17	3
Original reason for Medicare eligibility		
Old age and survivor's insurance	370	59
Disability insurance benefits	240	38
End-stage renal disease (ESRD)[a]	19	3
Hospice[b]	12	2
Medicare/Medicaid dual status, percent dual[b]	186	30
HCC score[c]		Statistic
Mean		2.24
25th percentile		0.85
Median		1.55
75th percentile		3.08

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which an individual was seen by a physician extender or treated in a CM medical unit. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Consistent with their poor health status, program participants had high rates of Medicare expenditures and service use in the year prior to enrollment—particularly in the two quarters immediately prior to participation. In Table III.3, we report baseline utilization and expenditure data for a common set of measures, including the four core measures from the Center for Medicare & Medicaid Innovation (CMMI). By avoiding costly ED use among low-acuity beneficiaries and connecting them to appropriate care, the Mesa Fire and Medical Department expects to reduce expenditures and ED utilization compared with what would have taken place absent its program. We examined baseline expenditures by calculating average per beneficiary per month (PBPM) Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $2,549. Average PBPM Medicare payments for inpatient care ($916), physician services ($619), and outpatient care ($331) were the largest drivers of payments—representing three-quarters of the total cost of care. Growth in expenditures beginning in the last two baseline quarters, particularly for inpatient care and physician services, appears to be an important factor in the higher total average PBPM Medicare expenditures in the period before enrollment.

Because the CCRI focuses on reducing ambulance transports to the ED, we also examined average PBPM Medicare payments on ambulance transports in the baseline period. We measured expenditures for (1) all ambulance transports and (2) ambulance transports most likely to be associated with a 911 call. Ambulance transports are identified in the claims data according to place of service. To identify the transports most likely to be associated with 911 calls, we received guidance from the Mesa Fire and Medical Department to focus on transports for which the final destination was a hospital and whose origin was (1) the caller's residence; (2) the scene of an accident or an acute event; or (3) a residential, domiciliary, or custodial facility such as a nursing home. Because the CCRI focuses on low-acuity individuals who call 911, we can expect the program to have the strongest effects on ambulance transports associated with these calls. Table III.3 shows that the average PBPM Medicare payments on all ambulance transports were $46 in the baseline year and $36 for transports most likely to be associated with 911 calls. Expenditures for ambulance use increased across the baseline quarters. The highest PBPM Medicare payments occurred in the quarter immediately before participation (a mean of $60 in the fourth quarter).

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Mesa Fire and Medical Department's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	629	577	599	627	629
Average Medicare expenditures PBPM[a]					
Total	2,549 (153)	2,177 (192)	2,279 (206)	2,568 (220)	3,105 (244)
Acute inpatient	916 (81)	741 (101)	799 (123)	951 (135)	1,149 (126)
Inpatient other[b]	307 (47)	254 (80)	184 (51)	362 (74)	415 (97)
Outpatient[c]	331 (27)	339 (40)	307 (36)	295 (27)	376 (32)
Physician services	619 (31)	527 (34)	574 (43)	612 (40)	750 (47)
Home health	122 (11)	93 (15)	103 (15)	129 (15)	161 (19)
Skilled nursing facility	185 (23)	164 (39)	251 (51)	162 (34)	165 (32)
Hospice	39 (15)	29 (15)	35 (15)	22 (11)	60 (18)
Durable medical equipment	30 (4)	30 (6)	26 (3)	34 (9)	30 (5)
Ambulance use (all transports)	46 (4)	34 (5)	41 (6)	46 (6)	60 (6)
Ambulance use likely associated with a 911 call	36 (4)	29 (4)	30 (5)	33 (4)	50 (6)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	975 (70)	791 (87)	826 (99)	999 (103)	1,231 (111)
Outpatient ED visits	2,158 (193)	1,729 (207)	1,796 (216)	1,927 (222)	3,113 (271)
Observation stays	339 (32)	315 (50)	225 (54)	346 (53)	459 (63)
Primary care visits in any setting	12,622 (605)	11,361 (688)	11,224 (767)	12,188 (752)	15,470 (1,092)
Primary care visits in ambulatory settings	7,815 (341)	7,539 (415)	7,530 (469)	7,753 (460)	8,370 (439)
Specialist visits in any setting	21,023 (988)	18,599 (1,246)	19,765 (1,318)	19,798 (1,158)	25,428 (1,563)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Specialist visits in ambulatory settings	11,060 (413)	10,815 (512)	10,719 (543)	10,719 (479)	11,808 (500)
Ambulance use (all transports)	1,519 (147)	1,204 (163)	1,297 (175)	1,476 (166)	2,054 (217)
Ambulance use likely associated with a 911 call	1,237 (131)	994 (144)	1,038 (162)	1,137 (138)	1,742 (195)
Measures of any health care utilization					
Percentage with a hospital admission[d]	43 (2)	15 (1)	14 (1)	18 (2)	22 (2)
Percentage with an outpatient ED visit[e]	58 (2)	22 (2)	22 (2)	25 (2)	34 (2)
Percentage with an observation stay[f]	24 (2)	7 (1)	5 (1)	8 (1)	10 (1)
Percentage with a 30-day readmission among all discharges	25 (2)	18 (4)	27 (4)	25 (4)	28 (4)
Percentage of participants with a readmission among all participants	11 (1)	3 (1)	4 (1)	4 (1)	4 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

As with expenditures, the average utilization of expensive Medicare services before enrollment was high, particularly in the last quarter before enrollment. The annual rate of acute care hospitalizations was 975 per 1,000 participants during the baseline year, with a rate of 1,231 per 1,000 beneficiaries during the last quarter. This compares to a national rate of 283 per 1,000 Medicare FFS beneficiaries in 2013. At baseline, participants had a high number of ED visits that did not lead to a hospitalization, at an annual rate of 2,158 visits per 1,000 participants in the baseline year, compared with the national rate of 445 per 1,000 Medicare beneficiaries in 2013. The highest rate was in the quarter immediately before participation, when there were 3,113 visits per 1,000 participants. At 58 percent, the likelihood of an ED visit in the baseline period was very high. Similarly, the 30-day unplanned readmission rate for participants was 25 percent per discharge—higher than the national rate of 18 percent per discharge. Because participants had high rates of acute care utilization, there may be a sizeable opportunity to reduce potentially avoidable admissions, ED visits, and readmissions during the post-period through the effective deployment of the CM units and by connecting beneficiaries with other forms of outpatient care. At the same time, Table III.3 also shows that participants had high rates of primary care utilization (7,815 primary care visits in ambulatory settings per 1,000 beneficiaries per year) and specialty care utilization (11,060 specialist visits in ambulatory settings per 1,000 beneficiaries per year).

We also examined the two measures of ambulance transport utilization. Over the baseline year, there were 1,237 ambulance transports per 1,000 beneficiaries that were likely associated with 911 calls and 1,519 overall ambulance transports per 1,000 beneficiaries. These rates were driven by much higher ambulance use in the last quarter before enrollment when there were 1,742 ambulance transports per 1,000 beneficiaries that were likely to be associated with 911 calls and 2,054 total ambulance transports per 1,000 beneficiaries.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

During the cooperative agreement, the awardee did not develop a new payment model, but instead pursued options under existing FFS payment, which covers 911 response for the medical calls using pre-hospital "treat/no transport" Medicaid billing codes, transport codes as appropriate, and codes for on-site behavioral health assessments.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

As the awardee has worked to determine feasible options for recouping the total cost of providing services, it has developed several iterations of its payment model. The final proposed payment model covers the following services: (1) on-site paramedic services to people who make low-acuity 911 calls and who do not require transport to the ED (treat/no transport), (2) transport to the ED for patients who cannot be treated at home, and (3) behavioral health clinician assessments of 911 callers with a behavioral health crisis.

The payment model is associated with pre-hospital billing codes in the 911 system as well as professional billing for a subset of the behavioral health services. The awardee's payment model does not cover the following program services provided under the cooperative agreement: (1) on-site nurse practitioner/physician assistant (NP/PA) services to low-acuity 911 callers who do not require transport to the ED, (2) transport to behavioral health facilities, (3) services provided by Mesa's captain paramedics on CM behavioral units, (4) dispatch of the units, and (5) care transition services to patients recently discharged from the hospital.

During the cooperative agreement, the flow of payments for medical services started with the awardee's vendor, Intermedix, which provides technology-enabled services including medical billing. Intermedix billed the payers on behalf of the awardee by using pre-hospital billing codes (transport and treat/no transport). Payers would reimburse for the services and Intermedix would send reimbursement to the awardee, which then directed payments to the CCRI fund. The behavioral health payments flowed in a similar way. CPR, the awardee's behavioral health partner that employed behavioral health clinicians, billed payers for reimbursable behavioral health services. The payers reimbursed CPR for services and CPR sent the payments to the awardee, which then added them to the CCRI fund.

C. Status of the payment model

The awardee had exploratory discussions with local payers and hospitals throughout the cooperative agreement, but had little success convincing these organizations to enter into payment model agreements that would help sustain the program. Although payers and hospitals saw value in the program, they generally were not motivated to enter into formal agreements,

citing competing priorities, uncertainty in the health care market, and uncertainty about a return on investment. The awardee almost entered into an agreement with a small Medicaid payer, who would have paid a $50 flat fee per patient encounter, well below projected cost. However, when the awardee decided to use the pre-hospital codes because it could get higher reimbursement, leaders decided not to enter the contract with the payer.

After the awardee began billing payers for a subset of services provided, the awardee tried to negotiate a shared savings agreement with payers and a hospital system, which did not materialize. Initially, one payer was interested but decided not to move forward, because current reimbursement for treat/no transport is less than what the awardee would need to cover clinician costs. As a fire and medical department, the awardee will arrive and treat patients who call 911 regardless of reimbursement, leaving payers with little incentive to partner. The awardee reported that payers would be more likely to reimburse at a higher rate if Medicare and Medicaid offered similar reimbursement for 911 community medicine services. The shared savings agreement with the hospital system would have allowed the awardee to use savings from transitional care services with their hospital partners for high risk patients discharged from the hospital before going to home health care. However, by the end of the cooperative agreement, the awardee decided to focus on its core business model of 911 services and not continue the hospital transition component during the no-cost extension.

D. Factors associated with the development of the payment model

The awardee said two conditions were critical to the development of its payment model: (1) access to pre-hospital billing codes through its Certificate of Necessity license with the state of Arizona, and (2) access to behavioral health billing codes through its partner Crisis Preparation and Recovery. Access to these billing codes, specifically for ambulance transport, paramedic treatment with no ambulance transport, and the services of behavioral health clinicians, enabled the awardee to recoup some of the cost of community medicine services.

In working to identify strategies that would allow the awardee to be reimbursed for the full cost of its program services, the Mesa Fire and Medical Department faced a number of obstacles. First, Medicare does not use treat/no transport billing codes or reimburse for those services, so the awardee could only use this code to bill Medicaid and some commercial payers. Second, transport and treat/no transport billing codes assume paramedics are the service providers. Consequently, compensation did not cover NP/PA services. Even if Medicare developed a billing code for treat/no transport that covered NPs and PAs, the awardee would still have challenges billing Medicare for the clinical services that NPs and PAs provide, because the awardee's tax identification number categorizes the Mesa Fire and Medical Department as providing ambulance services, not medical services, and municipalities cannot have two tax identification numbers. Conversely, the awardee's hospital partner, Mountain Vista Medical Center, is not authorized to bill payers for NP and PA pre-hospital services, which prevented the awardee from getting reimbursed through its hospital partner, as it did with behavioral services.

Finally, the reimbursement the awardee did get for behavioral health services did not fully cover the awardee's costs of providing the services. Medicare reimbursed for the assessment and crisis intervention counseling, but did not cover the costs for the services of Mesa's captain paramedics, the dispatch of the unit, or transportation of the patient to a behavioral health

facility. During the cooperative agreement, the awardee covered the shortfall with HCIA R2 funds and the Mesa Fire and Medical Department's operational budget. Going forward, however, this arrangement will not enable the awardee to sustain the volume of services it provided during the cooperative agreement.

The awardee talked to CMS about some of the challenges in developing a viable payment model given that payers lacked incentives to participate. The awardee believes the only plausible option is revised pre-hospital billing codes that reflect higher levels of reimbursement and staffing. The awardee is hopeful that CMS will develop new reimbursement codes for Medicare that can be adopted by other payers.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The Mesa Fire and Medical Department planned to partially sustain the CCRI program after the cooperative agreement ended. Specifically, it is maintaining the program's 911 response component through the CM units, but with fewer staff and reduced operating hours. The awardee hopes to sustain this component by continuing discussions with CMS about creating a billing code for community medicine that covers clinician costs. As noted, the awardee discontinued the care transitions component.

The cities of Los Angeles and Anaheim, California implemented programs like CCRI's 911 response component, and Anaheim's program was modeled after the awardee's pre-HCIA pilot community medicine program.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2 of the award, the Mesa Fire and Medical Department began implementing strategies to help sustain the program. The awardee was billing for behavioral health services

through its partner and was finalizing a payment contract with a small, local Medicaid managed care organization, although ultimately it did not contract with this payer.

In addition to negotiating potential payment models with other payers, the awardee was also expanding awareness of the program among paramedics by rotating program staff at all fire stations, not just the participating stations. The Mesa Fire and Medical Department also had several plans to scale the program, including expanding the 911 response component (after implementing a viable payment model). Other health systems reportedly expressed interest in using the program. The awardee was communicating with EMS agencies that replicated the program and others that were interested in replicating it, but not before they saw demonstrated program impacts.

C. Implementing the SSR plan: progress and changes

Sustainability. As discussed in Chapter IV on the payment model, in the third year of the program the awardee was struggling to secure payer participation. To sustain the program, the awardee was working to convince CMS to create a billing code for community medicine that covers clinician costs.

There were no strong prospects for sustaining the program, with either the awardee's own resources or funding from program partners. Although the program enjoyed the solid support of the city, Mesa's revenue base fell after a major city tax initiative failed, and the fire department's overall budget was cut by 5 percent. Efforts to convince the hospital partner to provide ongoing funding to staff the care transitions component stalled because there was no strong evidence that the program reduced hospital readmissions. The lack of funding led the awardee to phase out the care transitions component during Year 3. The awardee planned to use its six-month no-cost extension in part to focus on gaining support to sustain the 911 response function through the CM medical and behavioral units.

The awardee worked to sustain both the medical and behavioral health components of the program, but also tried to reduce costs to make the program more palatable to prospective funders. During Year 3, the CM medical units began operating at reduced hours—switching from a 24/7 model to a 40-hour per-week model. The awardee was in the process of determining which days and times are most important and effective, based on call volumes and other factors. The program downsized its staff, mainly through attrition when people left because they worried about their positions ending after the cooperative agreement. In the absence of mechanisms to cover the costs of the program, the awardee expected to reduce staff further, by about 80 percent (down to two medical practitioners and one behavioral health practitioner).

Scalability. Components of the program showed promise for scaling. The awardee had early discussions with at least one payer about adapting the community medicine program to target frequent ED users, and Dignity Health was potentially interested in adapting the care transitions model as part of its care coordination efforts.

Replicability. The awardee gave presentations about the program and reported high levels of interest from other fire and emergency medicine 911 response systems, both local systems beyond Mesa's jurisdiction and others throughout the country. They awardee found that some

payers would be more interested in funding a program if it were in a larger catchment area, like Phoenix. In 2015, Anaheim Fire and Rescue implemented a similar pilot program modeled after the awardee's program, and the City of Los Angeles Fire Department followed suit in January 2016. Another comparable program was implemented in Colorado, but it is unclear whether that program is a replication of Mesa's program.

D. Factors associated with progress toward implementing the SSR plan

Although program stakeholders valued and supported the program, particularly the 911 response component, the program faced several barriers to SSR. The main barriers were related to the challenges of developing the capacity to bill (through new billing codes and tax IDs that allow use of those codes) for services that are not currently reimbursed. The awardee also needed reimbursement from these codes at a level that covered the costs of the program. The awardee also needed support from hospital partners with enough of a financial stake in the care transitions component of the program to help sustain and potentially enhance it, as discussed in Chapter IV on the payment model. Also, respondents reported some staff issues that could affect SSR, including insufficient information sharing and other communication gaps between participating entities. Respondents said some of these issues stemmed from staff worries about whether their positions would last post-award.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension that allows it to continue to enroll new beneficiaries. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Mesa Fire and Medical department's program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicare beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Montefiore Medical Center

April 18, 2018

Jennifer Lyons (Mathematica Policy Research)
Danielle Chelminsky (Mathematica Policy Research)
Kate Stewart (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Montefiore Medical Center received a 12-month no-cost extension, which ends on August 31, 2018. The awardee ended its official enrollment period on August 31, 2017, but will continue to (1) enroll new patients and engage existing patients in the Behavioral Health Integration Program during the no-cost extension period, (2) test its payment model in the six participating sites, and (3) collect more data on the technology platforms that it developed.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

Montefiore Medical Center, a large tertiary care center in the Bronx, New York, is using its HCIA R2 funds to implement the Behavioral Health Integration Program (BHIP) in a subset of its 22 primary care sites (key program characteristics are noted in Table I.1). The BHIP centers on what the awardee calls a measurement-based collaborative care model, in which behavioral health staff in each primary care site use data from participants' initial screens and follow-up behavioral health scales to adjust their approaches to treatment (for example, psychiatric consultation and therapy).[4] Montefiore Medical Center rolled out the BHIP in two phases. The first phase was launched in February 2015, when three primary care sites began to implement the program. The second phase started in August 2015, when four additional sites began to implement the program. The awardee has since reduced its Phase 2 cohort to three sites, due to barriers to enrollment and lack of resources at the fourth site. Overall, Montefiore Medical Center expected the BHIP to reach 6,380 individuals, who receive services from participating primary care sites and who have depression, anxiety, or (for children and adolescents) attention deficit hyperactivity disorder (ADHD).[5]

Before the BHIP was implemented, most of the six primary care sites had some form of behavioral health care on site, and some sites were already implementing a measurement-based approach to care. In rolling out the BHIP, Montefiore Medical Center standardized the approach to care across the sites and added additional support for patient engagement through the creation of the patient educator role and the deployment of patient-facing technologies such as interactive voice response (IVR) and smartphone application platforms. The awardee hypothesized that these innovations will increase participants' satisfaction with care, improve their behavioral health and chronic physical health conditions, and reduce the costs of care as primary care providers and behavioral health staff work together on site to measure and respond to participants' behavioral health needs. (The awardee posited that alleviating an individual's depression and anxiety often leads to better self-care/self-management, better adherence to lifestyle changes and medications, and better physical health.)

[4] The BHIP is based on the Collaborative Care Model developed by the University of Washington Advancing Integrated Mental Health Solutions (AIMS) Center. See https://aims.uw.edu/collaborative-care.

[5] This report is based on data submitted by Montefiore Medical Center to the implementation and monitoring contractor. However, after submitting the data, the awardee reported that it planned to enroll and serve at least 4,725 direct participants. The awardee also reported that it planned to provide services to 4,725 indirect participants.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Montefiore Medical Center, a large tertiary care center in the Bronx, NY, is implementing the BHIP to provide behavioral health screening and treatment to eligible patients in 6 of its 22 primary care sites.
Major innovation	Implementing an existing integrated care model (the Collaborative Care Model) in a high-need population, and using technology platforms to improve patient engagement and service delivery
Program components	• Integrated behavioral health and primary care services • Patient registry to collect data on participants and monitor their progress through measurement-based care • Telemedicine tools to administer follow-up measures, and to engage and communicate with participants
Target population	6,380 individuals who receive services from participating primary care sites and screen positive for depression, anxiety, risk of alcohol use, or (for children and adolescents) ADHD
Theory of change/ theory of action	Primary care providers who work with on-site behavioral health staff to measure and respond to participants' progress will be better able to address participants' behavioral health needs. Improved access to behavioral health care (that is, integrated with primary care services) will lead to increased satisfaction with care, better physical and behavioral health outcomes, fewer hospitalizations, and lower costs.
Payment model	Value-based payments, monthly bundledpayment for care management and coordination services
Award amount	$5,583,090
Effective launch date	2/9/2015
Program setting	Six primary care practices and a virtual environment via telemedicine tools
Market area	Urban
Market location	The Bronx, NY
Target outcomes	• Increase in patient satisfaction • Improvement in participants' behavioral health and chronic disease outcomes • Net savings in the cost of care for the patient population through fewer hospitalizations and ED visits

ADHD = attention deficit hyperactivity disorder; ED = emergency department.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on six factors. First, the awardee enrolled 6,060 participants—95 percent of its enrollment target—by the end of the initial cooperative agreement. Second, the awardee delivered BHIP services to program participants as intended, as demonstrated by its self-reported service delivery metrics. Third, the awardee hired new staff and repurposed existing staff to provide behavioral health services through the program. Despite staff turnover, the awardee made steady progress in enrollment and service delivery. Fourth, the awardee engaged primary care providers in the BHIP and improved the level of engagement over the course of the program. Fifth, the awardee engaged participants in

program services through a combination of outreach by program staff and technology. Finally, participating clinicians and other implementation staff reported that the program had a positive effect on the delivery of care.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Montefiore Medical Center's BHIP, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. Montefiore Medical Center worked with two managed care organizations to develop a monthly bundled payment model to support the implementation of the BHIP. The managed care organizations provide coverage through Medicaid managed care, Medicare Advantage, and commercial plans. As of mid-September 2017, the awardee was in final negotiations with payers. It expected to implement the model at all six participating sitesin early 2018.

Sustainability plans. Montefiore Medical Center reported that it plans to use the no-cost extension period to (1) support the six participating primary care sites in their efforts to sustain the BHIP and (2) scale the program to its other primary care sites. The awardee intended to use its no-cost extension to further develop the payment model, engage site leaders, and test and improve program technology. Program leaders expected that the BHIP's alignment with the Delivery System Reform Incentive Payment (DSRIP) Program would make both implementing and non-implementing sites more interested in sustaining and scaling the program.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded around the start of the third program year with a sample of 30 potential respondents and achieved a response rate of 82 percent. The clinician survey was fielded in the second half of the third program year with a sample of 113 potential respondents and achieved a response rate of 71 percent.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

The awardee enrolled patients through a universal behavioral health screening protocol that included screening for depression, anxiety, and risk for alcohol use implemented at all participating sites. In this process, the front desk staff or sometimes the patient educator gave all patients who visited the BHIP's primary care practices a self-administered paper screening tool to complete while they were waiting to see their primary care provider. Patients identified their symptoms in the tool. A nurse scored the tool. If the patient screened positive, the nurse alerted the primary care provider, who discussed the BHIP with the patient. If the patient was interested, the primary care provider connected the patient to a member of the site's BHIP behavioral health team, which comprised of behavioral health patient educators, licensed clinical social workers, consulting psychiatrists, and—for children and adolescents—clinical psychologists. The process

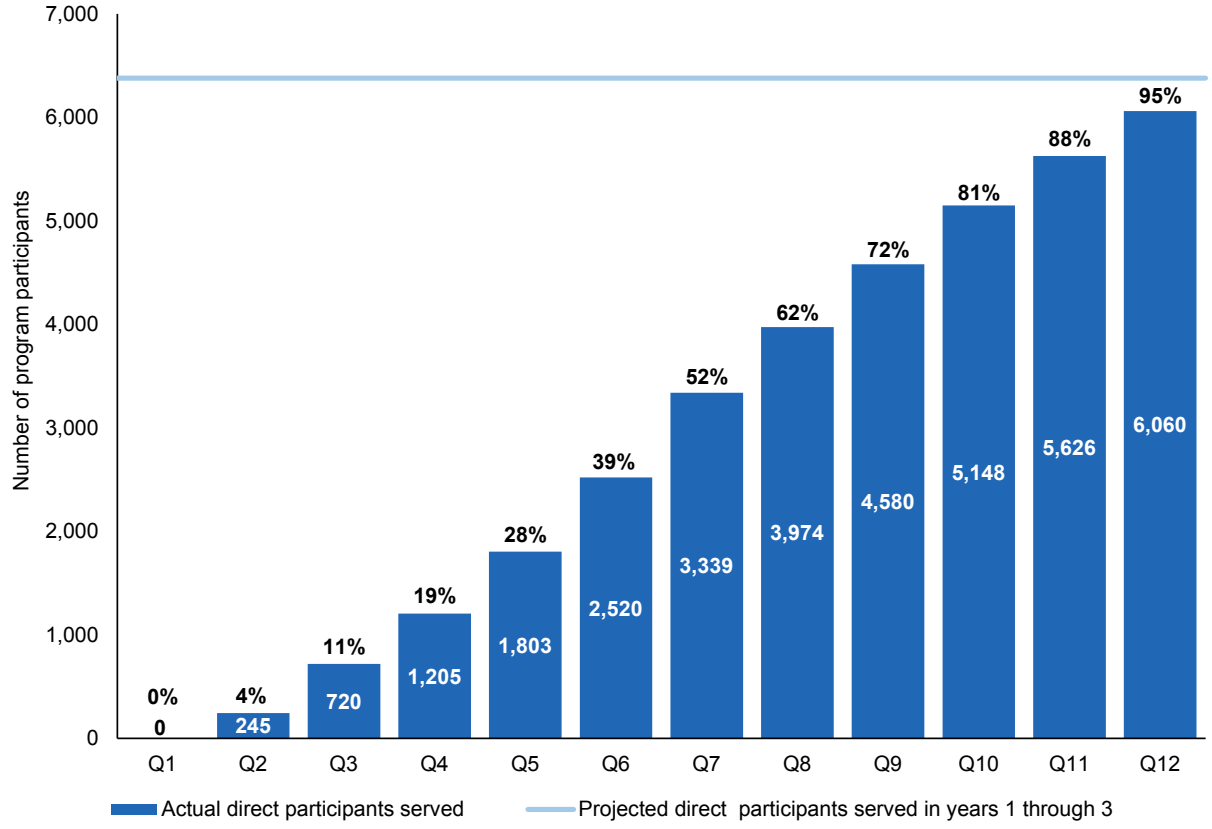
for connecting patients to BHIP staff varied by site and by provider. It could include in-person warm handoffs or the provider would follow up with the behavioral health staff by phone or through the electronic medical record (EMR) system. Patients who were interested in participating in the BHIP were then entered into the program's patient registry by the patient educator or social worker, indicating that they were enrolled in the program.

The awardee made one change to the eligibility criteria during the first program year: adapting its approach to enrolling participants with serious mental illness. Initially, program leaders intended sites to refer these participants to specialty mental health care rather than enrolling them in the program. However, given the potential for long wait times when referring participants to community mental health treatment, the BHIP shifted its strategy to engage these participants in short-term intervention at the primary care sites while seeking a long-term mental health treatment connection in the community. There were no major changes to the process for identifying and enrolling participants over the course of the cooperative agreement.

b. Evidence of enrollment effectiveness

Montefiore Medical Center achieved enrollment effectiveness. Overall, the awardee reported that it enrolled 6,060 direct participants from February 2015 (when it launched its program) through August 2017, which represents about 95 percent of its final three-year projection (Figure II.1). The awardee increased its projections from an original target of 4,575 direct participants to a projection of 6,380 direct participants beginning in Year 3 because it had almost met its original target by the end of Year 2 with 4,320 enrollees. The awardee reported 6,559 indirect participants in the 12th quarter, which was 89.9 percent of its projections.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee increased its projections from an original target of 4,575 direct participants to a projection of 6,380 direct participants beginning in Year 3.

c. Barriers and facilitators associated with enrollment effectiveness

Montefiore Medical Center's progress in meeting its three-year enrollment goal was influenced by several factors. Early in the cooperative agreement, the awardee encountered two enrollment setbacks that were resolved. In Year 1, the awardee struggled to maintain enrollment due to a disruptive rollout of a new EMR system. The awardee was able to overcome the issue by re-educating staff on the BHIP after the rollout and having staff actively engage patients in waiting rooms to administer the screen. In Year 2, the awardee ended the BHIP at one Phase 2 site that had struggled to identify potential participants. The awardee did not search for a replacement site because the other six participating sites were generating enough overall enrollment. Staff turnover at some sites became a barrier to enrollment when there were not enough staff available to take on new patients. The awardee reported some barriers to engaging participants that were identified through the enrollment process, such as language and literacy barriers, cultural stigma, and other patient concerns.

Despite these early barriers to the enrollment process (that is, the rollout of a new EMR system and difficulty enrolling participants at one site) and ongoing challenges related to staff turnover, Montefiore Medical Center succeeded in meeting its BHIP enrollment targets. The awardee noted that it increased the BHIP enrollment target in Year 3, which it attributed in part to the duration of treatment and enrollment being shorter than anticipated.[6] Program leaders said that they initially expected participants would stay in the program for 8 to 12 months (although anticipated time in treatment later decreased to 6 to 8 months after the start of the cooperative agreement), but participants typically needed only 3 to 4 months of BHIP engagement. As a result, the program was able to enroll more participants.

2. Delivery of program services

a. Description of and changes to service delivery model

The BHIP centered on the collaborative care model, a measurement-based model in which behavioral health care was integrated into primary care settings.[7] Under this model, all patients who visited the BHIP's primary care practices completed a self-administered screening tool in which they identified symptoms indicative of a behavioral health problem. Primary care providers connected patients who screened positive for one of the program's targeted conditions to a member of the site's behavioral health team (behavioral health patient educators, licensed clinical social workers, and—for children and adolescents—clinical psychologists). The process for connecting patients to the behavioral health team varied by site and provider but included in-person warm handoffs, phone calls, or communication through the EMR. Each site also had a psychiatrist who consulted with (1) the primary care providers to support the management of participants' psychiatric medications and (2) the behavioral health team to help identify participants' needs for behavioral health services. Depending upon their needs, participants were offered any combination of the following: short-term psychotherapy with the licensed clinical social worker, psychiatric medication management by the PCP with support from the psychiatrist, or telephone outreach with behavioral activation and symptom monitoring from the behavioral health patient educator. These staff monitored behavioral health measures and supported participants' efforts to achieve their behavioral health goals.

The behavioral health team conducted a follow-up by using the initial screen's full length corresponding measures (e.g., PHQ-9, GAD-7, AUDIT-C) to monitor participants' progress throughout the program. The screening targeted depression, anxiety, or risk for alcohol use based on symptoms identified in the participants' responses to the initial screening tool. The behavioral health team also used a patient registry, developed by the University of Washington, to collect and track patient screening scores, follow-up communications, and other related information. Participants whose scale scores did not improve during the program were connected by the patient educator or social worker to the consulting psychiatrist for further assessment and recommendations for next steps in treatment.

[6] In the final year of the program, the awardee increased its original enrollment goal of 4,575 participants to a new goal of 6,380 participants.

[7] University of Washington, Advancing Integrated Mental Health Solutions. "Collaborative Care." Available at https://aims.uw.edu/collaborative-care.

In the second and third years of the award period, the BHIP incorporated HIPAA-compliant telemedicine tools to boost participant engagement with the program. Participants could subscribe to an IVR technology that allowed them to complete follow-up monitoring measures via their phone, as well as to receive appointment reminders and simple health education messagges. BHIP staff also rolled out a smartphone application that provided participants with follow-up measurements; educational materials and videos; and reminders about treatment goals, appointments, and medications. The application also had a secure chat feature that allowed participants to communicate with patient educators. In addition, BHIP social workers could use a separate text messaging platform to receive and respond to messages from participants. Unlike the patient registry and other telemedicine tools, the text messaging platform was available to social workers in all of the awardee's primary care practices, not just those practices participating in the BHIP (or those who were part of the BHIP team).

b. Evidence of service delivery effectiveness

Overall, Montefiore Medical Center was successful in implementing the BHIP. Data collected through site visit interviews and surveys, as well as reported by the awardee, suggest that the awardee consistently met its targets for service delivery and participant engagement. Although the awardee faced some challenges with retaining staff and engaging primary care providers, program leaders worked with staff at the affected sites to identify strategies to overcome these barriers throughout the cooperative agreement.

Delivery of intervention services. The awardee was successful in delivering intervention services. In site visit interviews in the third year of the program, BHIP staff and primary care providers reported that the program achieved consistency in its approach to providing measurement-based behavioral health care to program participants. Both behavioral health team staff and primary care providers noted that they had a clear understanding of the program's workflow and expectations in the third year of the program.

Data reported by the awardee provided some support for these statements of service delivery success. One indicator was whether the program achieved its universal screening goal for all patients visiting the primary care site. Among all adult patients who visited the six primary care sites, the depression screening rate improved from 55 percent at the beginning of the program's second year (the first period for which screening data was reported) to a nearly 90 percent depression screening rate at the end of the third program year.[8] After participants were enrolled, the BHIP aimed to provide at least three follow-up visits within the first 10 weeks. The awardee reported that the program began meeting this goal at the end of the first year of program implementation, with participants receiving nearly four visits on average. By the end of the third

[8] The awardee did not report on progress for anxiety and alcohol use screening in its monitoring reports. There were no baseline rates for anxiety and alcohol use screening because those screens had not been done at the primary care sites prior to BHIP. However, since the depression, anxiety, and alcohol screens were all collected together as part of the BHIP universal screens, the awardee noted that the rates of anxiety and alcohol use screening after BHIP should be very similar to the rates of depression screening.

year, participants had about seven follow-up visits on average during their time in treatment, which averaged 13 weeks [9].

Another marker of intervention delivery success was the extent to which participants whose measurement results showed that their symptoms did not improve within the first three months of treatment were elevated to a higher level of care within the behavioral health team (that is, a consultation with the psychiatrist). According to awardee-reported data, by the end of the third program year, 94 percent of participants who had not improved within three months received a psychiatric consultation—up from 36 percent at the end of the program's first year. The program began meeting its target of at least 75 percent toward the beginning of the program's second year.

Staffing and training. The awardee was partially successful in meeting its staffing goals. The awardee faced some initial hiring delays during the first year of the program due to administrative procedures at the broader Montefiore Medical Group. However, these delays were resolved quickly and the awardee achieved its staffing target by the end of the first year of the program. Despite successful hiring, turnover remained a challenge throughout the program. BHIP leaders noted that turnover occurred across roles, although respondents in the site visit interviews noted particular challenges with the social worker and psychiatrist roles. This challenge was reflected in the data reported by the awardee. In the third program year, the awardee reported that nine social workers, two psychiatrists, and one patient educator left their positions. In addition, the program needed to fill temporary staff absences (for example, for jury duty and maternity leave), which exacerbated the challenge.

The awardee succeeded in meeting its training goals. Montefiore Medical Center provided training to staff that appears to have contributed to the service delivery effectiveness. The awardee consistently outpaced its training targets throughout the program. Site visit interview and survey findings suggested that program staff and clinicians engaged in and made use of these trainings. In our survey of non-clinician staff, 96 percent of respondents said they attended formal trainings and 73 percent said they learned new skills that were important for their role. Moreover, 81 percent of non-clinician respondents said that the training helped them improve their job performance. Outside of formal program trainings, all non-clinician survey respondents said they went to staff meetings, 81 percent said they were trained through individual supervision, and 76 percent through group supervision. In addition, 90 percent of non-clinician staff received support by asking a colleague for help and 71 percent received technical assistance. Only 27 percent of staff either somewhat or strongly agreed that other training would have been helpful for their responsibilities with the program.

Engagement of providers. Montefiore Medical Center succeeded in engaging primary care providers in the BHIP. Because all primary care providers within a site were required to participate, no recruitment was necessary. In the site visit interviews, BHIP leaders and behavioral health staff noted that most primary care providers expressed interest in and support for the program. However, in the first two years of the program, these staff also reported that many providers had trouble making time for the program and coordinating with BHIP staff on

[9] The awardee did not specify in its reporting whether the average number of follow-up visits occurred within 10 weeks or rather were averaged across the duration of participants' treatment.

participants' care, given the many other tasks and expectations placed on them. During site visit interviews, primary care providers confirmed both this supportive attitude and the concerns about their availability to engage in the program. By the third year, BHIP staff noted that primary care providers were connecting more seamlessly with the program, in part due to improved workflow processes between primary care providers and the behavioral health team and adaptations to individual provider's communication styles. Although primary care providers continued to report competing priorities vying for their time, they also said that they were working well with the behavioral health team and found value in BHIP services.

Our survey findings identified similar themes related to provider engagement. In the clinician survey (which included primary care providers as well as other types of clinicians such as BHIP consulting psychiatrists), 79 percent of clinicians said they were very satisfied or somewhat satisfied with their current role in the BHIP. Nearly three-quarters of clinicians (72 percent) reported that the program made their job easier and more than two-thirds (69 percent) said that their ability to provide care to their patients improved under the BHIP. When asked why they wanted to participate in the program, clinicians overwhelmingly indicated (93 percent) that they thought the program would improve access to care or improve care for their patients. Clinicians indicated that the main reasons they would not want to participate in the program were lack of appropriate staffing resources (61 percent), too much of a time commitment (46 percent), or too many requirements (46 percent).

One way BHIP leaders and behavioral health staff engaged primary care providers was through staff and team meetings in which the behavioral health team presented data on a specific participant's progress in the program and sought consultation on strategies for improvement. In the clinician survey, 88 percent of clinicians said they attended staff meetings and 70 percent said they attended huddles. In addition, nearly two-thirds (66 percent) of clinicians said they attended formal BHIP training. Of those, two-thirds (67 percent) strongly or somewhat agreed that the trainings helped their job performance.

Engagement of program participants. The awardee was successful in engaging participants. In site visit interviews, BHIP leaders, primary care providers, and behavioral health staff noted that the program consistently and successfully engaged participants in services. This sentiment was reflected in the survey findings, with the majority of clinicians (81 percent) and program staff (91 percent) agreeing that BHIP succeeded in engaging participants. Monitoring data submitted by the awardee also provided evidence of successful engagement of participants in the program. Recognizing that not all participants identified as needing behavioral health care would be interested in treatment, BHIP staff aimed to engage at least three-quarters of those patients who screened positive on their initial assessment. The awardee reported that the program consistently exceeded this goal, starting at the beginning of the second year of the program, and ultimately achieved an 89 percent engagement rate among participants identified as in need of behavioral health care. The BHIP also succeeded in engaging participants in a timely way and according to their goals. The program aimed to engage participants in care within two to four weeks of the initial screening. At the end of the third year of the program, the awardee reported that participants' first follow-up appointments with the behavioral health team occurred 18 days after the screening, on average. This is an improvement over the 26-day average reported toward the end of the program's first year of implementation.

c. Barriers and facilitators associated with service delivery effectiveness

Delivery of intervention services. In site visit interviews, BHIP staff reported that the program's adaptability, technology, and workflow improvements facilitated successful implementation of the program. In each year of program implementation, site visit interview respondents noted the importance of having the flexibility to tailor the program model to their primary care site's unique needs and culture. In particular, both behavioral health and primary care staff appreciated the ability to tailor the communication and workflow processes to the preferences, availability, and existing communication styles of the primary care staff. In addition to the program's overall design, BHIP behavioral health staff pointed to the program's technological tools (IVR and smartphone application) as important improvements in the final year of the program. In particular, social workers and case managers noted that these tools allowed them to do their jobs more efficiently, by providing automated means for collecting follow-up screening data and communicating with participants as well as by improving their success rates in making contact with participants (for example, via text message or smartphone application chat function). Staff reported that by reducing the need for routine follow-up (in particular, to collect data on monitoring measures), these technologies gave them time to tackle follow-up communication for participants identified through monitoring as having more immediate needs. Finally, both primary care providers and program staff noted in interviews that they improved their workflow and communication over the course of the program—for example, engaging in more warm handoffs by the third year of the program and more consistently and efficiently reviewing patients' initial screening results to prioritize patients with the highest scores for the soonest available appointments.

Overall, BHIP respondents in site visit interviews and surveys identified few barriers to service delivery, outside of the turnover challenges. Limited staff time was the most commonly cited barrier to the program achieving its goals, with nearly 80 percent of non-clinician staff and approximately 70 percent of clinicians who responded to the survey noting this as a barrier. However, most of the respondents in both groups described this barrier as minor. More than half of non-clinician program staff (56 percent) cited documentation requirements as a barrier to the program achieving its goals, although staff did not report challenges with this during site visit interviews and a much smaller proportion of clinicians (35 percent) noted this as a challenge. Despite the appreciation of program flexibility noted in site visit interviews, half of the BHIP staff who responded to the survey noted that integrating the program into existing structures posed a barrier to achieving the program's goals. However, only 14 percent of respondents described this as a major barrier.

Staffing and training. As described previously, Montefiore Medical Center succeeded in hiring new staff and engaging existing staff to support its program implementation efforts. However, the awardee struggled with staff turnover over the course of the cooperative agreement. BHIP leaders noted that the program lost some staff due to the program serving as a launching pad for their careers—for example, social workers leveraged the program's innovative training in delivering integrated care to seek higher paid positions elsewhere. Overall, most site visit interview respondents noted that they did not believe turnover to be associated with the program, but rather as relatively typical for social workers and psychiatrists, among whom the most turnover occurred.

In site visit interviews in the third program year, some staff noted that turnover posed challenges for managing their own workload and maintaining participant engagement in the program. Some respondents noted that staff turnover was particularly challenging given the nature of behavioral health work, with one respondent commenting that participants can "lose interest" after a staff member with whom they developed a rapport leaves the program. This challenge was reflected in our survey findings. When asked about barriers to the BHIP achieving its goals, more than three-quarters of non-clinician staff and approximately two-thirds of clinicians identified staff turnover and unfilled positions as a barrier. However, behavioral health and primary care staff noted that the program identified strategies to mitigate these challenges. For example, the program sometimes reallocated existing staffing resources across sites to fill in the gaps. BHIP leaders and staff also reported using the program's technology to fill the engagement gaps while sites were understaffed. Respondents reported that teams were able to maintain contact with patients using the IVR and smartphone application tools to stay connected and keep patients engaged between appointments. In particular, patient educators noted that this technology helped them manage their communications and made their workload more manageable.

Site visit respondents also noted that ongoing training from program leaders had been a major facilitator to the BHIP program's success. Over the course of the program, BHIP leaders provided trainings and retrainings on a variety of topics as needed, such as on BHIP protocols and goals, evidence-based therapy interventions, workflows, and technical assistance with the tele-health platforms. In addition to formal trainings, BHIP leaders held meetings with the behavioral health teams at each site, as well as role-specific meetings that convened staff in the same roles across the different participating primary care sites—for example, a regular convening of patient educators from different sites to discuss their challenges, share their techniques, optimize their role, and brainstorm new ideas for managing their workload. Staff in all roles agreed that these trainings provided them with support throughout the program.

Provider engagement. All staff reported that primary care providers had limited time to engage with the program due to their many other responsibilities and that this limited time sometimes posed a barrier to provider engagement. BHIP leaders and staff noted that over the course of the program they developed different strategies for engaging providers in working with the BHIP behavioral health team, which helped mitigate this challenge. One strategy was to tailor their communication approach to individual primary care staff (for example, some primary care providers preferred to conduct warm handoffs in person, while others preferred to communicate by phone or through the EMR messaging system). Another strategy was to educate the primary care providers on the program and its benefits. In the second and third years of site visit interviews, both BHIP behavioral health staff and primary care providers reported that seeing evidence of participants' progress—in the form of improvement on the follow-up measures—increased provider engagement in the program.

> "One thing that's helped with buy-in is because the doctors are seeing that not only are there patient depression and anxiety screens, the symptoms are coming down, but that they're losing weight or their A1c is lowering. So, I think that overall, you know, the doctors are seeing these changes. And I think that's a testament that the program is working."
>
> —Social worker

Despite these successes, some respondents noted that a minority of primary care providers remained reluctant to prescribe and manage psychotropic medications, instead relying on the psychiatrist to address the behavioral health prescribing needs. The program worked to address this challenge by providing additional training and consultation to primary care providers. Respondents noted that these efforts helped increase their comfort with prescribing psychotropic medications appropriate for primary care practice.

Participant engagement. Over the course of the cooperative agreement, BHIP staff reported that participants were generally receptive to the program and noted that few participants declined to participate. In interviews, some respondents suggested that participant engagement improved over the course of the program as primary care providers and the behavioral health team became more comfortable "selling" the program's purpose and services to participants. In addition, BHIP staff and primary care providers noted that the BHIP serves a large number of participants in need of what they called "concrete services," such as housing, food, and transportation assistance. Staff felt that the program was well positioned to meet these needs, both directly through BHIP staff and through referrals to Montefiore Medical Center's community health workers, who supported some BHIP staff in connecting participants to outside social services. Staff reported that the program's ability to connect participants to this support helped engage participants in other aspects of their care. In the third year of the program, nearly all staff who participated in site visit interviews identified the program's telemedicine tools as critical to engaging participants in the BHIP. Staff reported that these technologies improved participant engagement by allowing participants to engage with the program on their own schedule. For example, the smartphone application allowed participants to complete the program's follow-up monitoring measures at any time of day or night. The measures allowed the behavioral health team to monitor participants' symptoms and progress and pursue additional follow-up communication if the measures indicated that symptoms were worsening or not improving.

Although most site visit respondents noted that participants were receptive to the program and appreciated the different approaches to engaging with it, staff reported some barriers to engaging participants in care. When asked the reasons participants declined to participate in the program, roughly three-quarters of BHIP staff named cultural stigma and about two-thirds cited already being linked to an outside behavioral health provider, patient concerns about financial burden or insurance coverage, and patients not perceiving their behavioral health needs as serious enough to require care. Approximately half of the clinicians and staff responding to the survey noted that participants' literacy and language posed barriers to engaging them in care. This finding was reflected in site visit interviews across all years of the program, with respondents noting that the program serves a diverse population with many different language needs. The program attempted to address this challenge by employing bilingual staff and engaging translation services whenever possible, but some staff noted that these factors continued to pose a barrier.

C. Assessment of perceived program effects on the delivery of care and outcomes

According to the survey, the majority of staff perceived that the BHIP program had positive impacts on the delivery of care, access to services, and participant satisfaction and quality of life

(Table II.2). For example, all respondents indicated that the program had a positive impact on care coordination and more than 90 percent noted that the program improved quality and efficiency of care, fair provision of care, quality of life, and achievement of participants' health goals. Clinicians who participated in the program and responded to the survey also endorsed the BHIP's positive effects, although at slightly lower rates.

Table II.2. Staff and clinician perceptions of BHIP program effects on care

Percentages of staff indicating that the BHIP program had a positive impact on the following:	Number of non-clinician respondents (N = 23)	Percentage of non-clinician respondents	Number of clinician respondents (N = 65)	Percentage of clinician respondents
The quality of care and services you provide to participants	21	91	54	81
Your ability to respond in a timely way to participant needs	19	83	49	75
The efficiency of care services provided to participants	21	91	n.a.	n.a.
Your ability to provide care or services that are responsive to participant preferences, needs, and values	16	70	48	73
Care services that are provided fairly to all participants	21	91	n.a.	n.a.
Access to care or services for all participants	20	87	52	81
Achievement of participants' health goals	21	91	49	74
Participant satisfaction	20	87	50	77
Participant quality of life	22	96	50	77
Care coordination	23	100	51	78

Source: HCIA R2 evaluation survey of participating non-clinician staff (fall 2016) and HCIA R2 evaluation survey of participating clinicians (spring 2017).

n.a. = not applicable.

In addition, nearly all BHIP non-clinician staff who responded to the survey (96 percent) felt that the program was either very or somewhat effective at achieving its goals; the remaining 4 percent said it was too soon to tell. Similarly, most clinician respondents (88 percent) agreed that the program was very or somewhat effective in achieving its goals, although a small proportion (12 percent) responded that the program was somewhat ineffective.

> "I think we're seeing people that would never in a million years have ever seen a psychiatrist. I think [patients] are feeling more comfortable accessing their behavioral health care under the same roof as their PCP. I think they feel very reassured that we're talking with their PCP."
>
> —Psychiatrist

In interviews, the program's behavioral health and primary care staff echoed the view that the program had positively affected quality of care and care coordination and, as a result, participants' health outcomes and quality of life. Staff reported that communication between primary care and behavioral health staff improved over the course of the program. In addition, primary care providers noted that the program provided them (and their patients) with behavioral health resources that would have otherwise been difficult to access. As one primary care provider

put it, "This work is time-intensive, people-intensive, hand-holding-intensive, and we don't have the time and the expertise to do that. So to push it into the social work realm and to use the psychiatrists and the whole team—it makes it a lot easier for us."

> "Overall, it just makes the work more meaningful. I think you just feel like you're delivering a more holistic care and overall care for the patient. You're able to meet all their needs and not just think in a bubble."
>
> —Primary care provider

Moreover, both behavioral health staff and primary care providers noted that the program's positive effects on participants helped increase provider engagement. In addition to improving participants' behavioral health symptoms, some respondents noted that they also saw improvements in participants'' physical health (for example, improvements in weight, blood pressure, blood sugar, and lipids). Several staff noted that they expected the program's focus on both physical and behavioral health care to result in gains in both areas.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Given that Montefiore Medical Center met its targets for enrolling and engaging participants in a timely manner and provided services as planned, we should be well positioned to identify program outcomes. Further, qualitative information from awardee staff indicated that they observed improvements in patients' health status. We expect such improvements to lead to fewer outpatient ED visits, inpatient admissions, and unplanned readmissions, as well as lower health care spending. However, program leaders noted that some expected BHIP outcomes may not appear until after our evaluation period has ended. For example, program leaders noted that integrated care initiatives often do not see cost savings until two to four years out. If this is the case for the BHIP, regardless of the comparison group selected, we may likely only observe cost savings for participants enrolled and served early in the award period.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Montefiore Medical Center's BHIP and the baseline demographic characteristics as well as the cost and utilization characteristics of Medicare fee-for-service (FFS), Medicaid FFS, and Medicaid managed care beneficiaries who were enrolled in the BHIP. For Medicare beneficiaries, the period of analysis ran from the program launch date on February 9, 2015, through May 31, 2016. For Medicaid beneficiaries, the period of analysis ran from February 9, 2015, through June 30, 2015. The period of analysis for Medicaid beneficiaries was shorter because New York State provided data for Medicaid services through June 30, 2015, only and more recent data is not yet available. For all Medicare and Medicaid beneficiaries, the baseline year comprised the 365 days before each participant's enrollment date. A total of 609 BHIP participants—122 Medicare beneficiaries and 487 Medicaid beneficiaries—were included in this analysis.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Montefiore Medical Center

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	175[a]
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018	4,555[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	3,878
Likelihood of all-cause hospitalizations	2,932
MDE sample size requirement to detect 20% effect	
Total expenditures	970
Likelihood of all-cause hospitalizations	733
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline

Table III.1 *(continued)*

Evaluability domain	Response
Claims sufficient to identify treatment and comparable comparison group?	Questionable, moderate rate of identification of treatment group with claims with some dilution of treatment effect within an intent-to-treat framework
Likelihood of solid comparison group	Some issues, but probably surmountable; expect to select a comparison group
Do claims identify the primary expected effects?	Some effects observed in claims data, but important effects likely missing
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician, staff, and beneficiary surveys
Implementation data that will be analyzed	Possible electronic health record data

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We anticipate conducting a rigorous impact analysis for Medicaid beneficiaries in the BHIP. We will use propensity score matching techniques to select comparison beneficiaries who live in the Bronx, have one of the program's target conditions, and who are similar to BHIP participants in terms of key characteristics. Because provider judgment is used in enrolling beneficiaries, we will explore including all patients with the target conditions who have visited a BHIP physician as part of the treatment group (rather than only including enrolled patients). When enrollment ended (August 2017), there was a sufficient number of enrolled Medicaid beneficiaries to identify plausible effects on claims-based outcomes (see Table III.1).

B. Characteristics of Medicare FFS beneficiaries at baseline

Montefiore Medical Center began to enroll Medicare and Medicaid beneficiaries in the BHIP on February 9, 2015. By the end of November 2016, the awardee had enrolled 4,679 participants, most of whom were Medicaid beneficiaries (see Section C). We linked 1,091 participants to the Medicare enrollment database; 160 participants met the Medicare FFS enrollment criteria.[10]

In presenting the baseline characteristics, we restricted the treatment group to Medicare FFS beneficiaries, both Parts A and B, who had Medicare as the primary payer when they became eligible for awardee-provided services (that is, their enrollment date) and who met all program criteria for a period of 90 days during the 365 days immediately before their enrollment, including their date of enrollment. In addition, they had to be enrolled in the BHIP on or before

[10] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

May 31, 2016, in order to ensure that they would be exposed to the intervention for long enough to capture nearly all claims for the most recent participants. The calendar period covered by the baseline quarters varied by participant because it was based on each participant's enrollment date. The baseline characteristics presented for this report are as of May 2016, at which point the BHIP had a total of 3,139 participants. As noted above (and described in Section D below), the majority of these BHIP participants were enrolled in Medicaid. We were able to link 792 participants to Medicare data. Most of these Medicare beneficiaries were enrolled in managed care and were not included in this analysis because the data for these beneficiaries were not available. After excluding individuals who did not meet the above criteria, we were able to obtain baseline characteristics for a total of 122 participants, or 15 percent of the beneficiaries who were linked to Medicare data.

The Medicare FFS beneficiaries participating in the BHIP during the first four program quarters were a diverse group of individuals in terms of demographic and health status characteristics (Table III.2). Forty-three percent of these participants were under the age of 65, whereas 7 percent were 85 or older. They were also far more likely to be female (78 percent) or to be part of a racial or ethnic minority group (66 percent). Black people accounted for almost half of the Medicare FFS beneficiaries alone (48 percent), which reflected the racial composition of the Bronx, where the intervention hospitals were located. Thirty-eight percent were dually eligible for Medicare and Medicaid, compared to only 18 percent nationwide. The percentage of beneficiaries who were enrolled in Medicare FFS because of disability (51 percent) was much higher than the national average of 24 percent. In addition, the average hierarchical condition category (HCC) risk score of 1.47 was 47 percent higher among participants who were Medicare FFS beneficiaries than the national average of 1. The distribution of HCC risk scores was associated with very high average expected expenditures. Taken together, the characteristics of Medicare FFS beneficiaries participating in the BHIP suggested that they were in poor health and that they had a great need for health care services.

Medicare FFS beneficiaries participating in the BHIP had high rates of service use and expenditures in the 365 days before enrollment, which was consistent with their needs. Table III.3 shows the baseline cost of care in total and by major types of services—calculated as the average Medicare payment per beneficiary per month (PBPM).[11] The total average Medicare payment PBPM during the baseline year was $1,754. Quarterly estimates ranged from as low as $1,563 in the third quarter to as high as $2,009 in the first quarter. The largest drivers of the total cost of care were the average Medicare payment PBPM for acute inpatient ($758), outpatient ($383), and physician ($318) services. These costs represented 83 percent of the total cost of care. Montefiore Medical Center expected to reduce hospitalizations and ED visits compared with what the rates would have been absent the BHIP by offering more comprehensive follow-up and preventive primary care than Medicare FFS beneficiaries typically received. If reductions in the use of these high-cost services were achieved, then a decline in Medicare expenditures for BHIP participants in the post-intervention period should become evident and would be reflected in our impact estimates.

[11] Months referred to in our calculations are 30-day periods rather than calendar months.

The average rate of acute hospital admissions was 497 per 1,000 Medicare FFS beneficiaries participating in the BHIP per year during the baseline year—approximately double the national average of 274 per 1,000 Medicare FFS beneficiaries per year.[12] Twenty-five percent of Medicare FFS participants were admitted to a hospital and 13 percent of those who were discharged had a readmission within 30 days, which was lower than the national average of 18 percent. The average annual rate of ED visits that did not lead to a hospitalization was 829 per 1,000 Medicare FFS beneficiaries participating in the BHIP—almost double the national annual rate of 445 per 1,000 Medicare FFS beneficiaries. Almost half (47 percent) of these participants had an outpatient ED visit. These figures suggested that there was an opportunity to reduce ED visits through the health care services provided by the BHIP. During the baseline year, the average annual rate of primary care visits in any setting was 5,843 per 1,000 participating Medicare FFS beneficiaries, whereas the average annual rate of primary care visits in ambulatory settings was 4,326 per 1,000 participating Medicare FFS beneficiaries. The high rate of specialist service use in ambulatory settings was 11,399 per 1,000 participating Medicare FFS beneficiaries, which suggested that there may be a need for better access to primary care—a key component of the BHIP.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Montefiore Medical Center's program through May 31, 2016

	All participants (N = 122)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	53	43
65 to 74	41	34
75 to 84	19	16
85 and older	9	7
Gender		
Female	95	78
Male	27	22
Race		
White	42	34
Black	59	48
Other	5	4
Hispanic	16	13
Original reason for Medicare eligibility		
Old age and survivor's insurance	59	48
Disability insurance benefits	62	51

[12] All national data in this paragraph are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

Table III.2 *(continued)*

Characteristics	All participants (N = 122)	
	Number	Percentage
End-stage renal disease (ESRD)[a]	1	1
Other status		
Medicare/Medicaid dual status[b]	46	38
HCC score[c]		**Value**
Mean		1.47
25th percentile		0.71
Median		1.12
75th percentile		1.88

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is the date when the participant became eligible for awardee-provided services. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3. Baseline year costs and health care utilization for Medicare FFS beneficiaries enrolled in Montefiore Medical Center's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	122	112	116	120	122
Average Medicare costs PBPM[a]					
Total	1,754 (331)	2,009 (564)	1,652 (371)	1,563 (392)	1,801 (455)
Acute inpatient	758 (215)	610 (217)	804 (271)	712 (291)	892 (331)
Inpatient other[b]	36 (33)	142 (140)	8 (8)	0 (0)	0 (0)
Outpatient[c]	383 (71)	438 (170)	342 (80)	368 (81)	387 (80)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Physician services	318 (39)	291 (48)	331 (49)	324 (56)	324 (48)
Home health	95 (28)	98 (45)	131 (44)	54 (28)	100 (41)
Skilled nursing facility	147 (50)	403 (165)	19 (19)	94 (66)	85 (84)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	16 (4)	27 (12)	17 (10)	10 (4)	12 (4)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	497 (105)	328 (104)	531 (152)	483 (153)	634 (162)
Outpatient ED visits	829 (161)	765 (199)	849 (201)	725 (164)	967 (220)
Observation stays	35 (35)	0 (0)	35 (35)	35 (34)	67 (46)
Primary care visits in any setting	5,843 (716)	5,570 (1,556)	5,874 (979)	5,386 (924)	6,503 (972)
Primary care visits in ambulatory settings	4,326 (478)	4,187 (1,403)	4,388 (824)	4,281 (744)	4,435 (378)
Specialist visits in any setting	14,774 (1,613)	14,963 (1,914)	14,120 (1,664)	12,912 (1,441)	17,008 (2,951)
Specialist visits in ambulatory settings	11,399 (1,083)	12,306 (1,458)	10,970 (1,211)	10,289 (1,127)	12,039 (1,369)
Measures of any health care utilization					
Percentage with a hospital admission[d]	25 (4)	8 (3)	11 (3)	9 (3)	13 (3)
Percentage with an outpatient ED visit[e]	47 (5)	15 (3)	18 (3)	16 (3)	18 (4)
Percentage with an observation stay[f]	3 (2)	0 (0)	1 (1)	1 (1)	2 (1)
Percentage with a 30-day readmission among all discharges	13 (5)	0 (0)	7 (7)	25 (13)	13 (9)
Percentage of participants with a readmission among all participants	3 (2)	0 (0)	1 (1)	3 (1)	2 (1)

Source: Mathematica analysis of information from the awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the

Table III.3 *(continued)*

91 days before the fourth baseline quarter, and so forth. The first baseline quarter will be 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and each baseline quarter according to the number of days in which each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

C. Characteristics of Medicaid FFS and managed care beneficiaries at baseline

In presenting the baseline characteristics of Medicaid beneficiaries participating in the BHIP, we restricted the treatment group to Medicaid FFS and managed care beneficiaries who were enrolled in Medicaid when they became eligible for awardee-provided services (that is, their enrollment date) and who met all program criteria for a period of 90 days during the 365 days immediately before their enrollment. After excluding individuals who did not meet the above criteria, we were able to link 2,210 participants to the New York State Medicaid data.

Although there were 3,408 Medicaid beneficiaries participating by November 2016, beneficiaries had to be enrolled in the BHIP on or before June 30, 2015, to be included in this report so that we had a full year of baseline Medicaid data. (We will re-run the analysis of baseline characteristics on all enrollees when more recent data for New York state become available). This criterion excluded a significant portion of Medicaid beneficiaries who enrolled in the BHIP after that date. We can report baseline demographic and health status characteristics for 487 participants (22 percent of those linked to the New York State Medicaid data). We can also report health care cost and utilization information for 455 participants (21 percent). Ninety of these participants were dual beneficiaries and 365 of them were non-dual beneficiaries. Health care cost and utilization characteristics are presented separately for dual and non-dual Medicaid beneficiaries. Based on the ICD-9 codes for Medicaid claims occurring during each beneficiary's baseline year, we used the Chronic Disability Payment System (CDPS) software to generate chronic condition categories and risk scores.

Table III.4 illustrates the demographic and health status diversity of Medicaid beneficiaries participating in the BHIP as of June 30, 2015. About 51 percent were middle-age adults (ages 35 to 64), whereas only 10 percent were children or adolescents (ages 3 to 17) and 15 percent were adults who were 65 or older. They were also far more likely to be female (78 percent) and either black or Hispanic (71 percent), reflecting the racial and ethnic diversity of the Bronx, where the intervention hospitals were located. Seventy-eight percent were not dually eligible for Medicare and Medicaid, 73 percent were covered by a managed care plan, and 97 percent received full Medicaid benefits. These participants became eligible for Medicaid for three main reasons: (1)

they received Temporary Assistance for Needy Families because they were low-income adults (32 percent), (2) they were eligible for Medicaid when it was expanded through the Affordable Care Act (23 percent), and (3) they received Supplemental Security Income because they were blind or disabled (21 percent).

BHIP Medicaid participants also had a wide range of chronic conditions, poor health status, and a great need for care. About 82 percent of them appeared in one or more CDPS categories. Their average CDPS risk score was 1.86—86 percent higher than the national average of 1 (Table III.5). Their chronic conditions included cardiovascular (37 percent), pulmonary (28 percent), skeletal and connective tissue (27 percent), and gastrointestinal (24 percent) conditions. Psychiatric conditions, however, were the most common chronic conditions by far: 53 percent of participants manifested these conditions. The distribution of chronic conditions reflected the fact that Montefiore Medical Center deliberately targeted individuals who had depression, anxiety, or an alcohol use disorder, as well as children and adolescents with ADHD. We also looked at whether participants had a Medicaid claim for any of these targeted conditions during the year before they enrolled in the program. We found that 80 percent had at least one Medicaid claim with a diagnosis for any of the four targeted conditions during the baseline year. In particular, 66 percent had at least one claim with a diagnosis of depression, 34 percent had a claim with a diagnosis of anxiety, 4 percent had a claim with a diagnosis of an alcohol use disorder, and 2 percent had a claim with a diagnosis of ADHD.

Table III.4. Baseline demographic characteristics of Medicaid FFS and managed care beneficiaries enrolled in Montefiore Medical Center's program through June 30, 2015

	All enrollees (N = 487)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
3–17 years (children and adolescents)	49	9
18–34 years (young adults)	116	23
35–64 years (middle-age adults)	248	50
65 years or older (older adults)	74	15
Gender		
Female	382	78
Male	105	21
Race/Ethnicity		
Non-Hispanic white	29	5
Non-Hispanic black	118	24
Hispanic	226	46
Other	11	2
Type of benefits		
Full Medicaid benefits	471	96
Restricted benefits	16	3

Table III.4 *(continued)*

Characteristics	All enrollees (N = 487)	
	Number	Percentage
Medicaid eligibility category		
SSI aged	27	6
Non-SSI aged	30	6
SSI blind/disabled	102	21
Non-SSI blind/disabled	14	3
TANF, safety net, or low-income family adults	155	32
Other adults	113	23
TANF, safety net, or low-income family children	25	5
Other children	21	4
Managed care enrollment		
Comprehensive managed care plan	356	73
Long-term care carve out	41	8
No managed care enrollment	90	18
Medicare/Medicaid dual status		
Dual	107	21
Non-dual	380	78
HCBS waiver enrollment		
Enrolled in any HCBS waiver	5	1
Not enrolled in an HCBS waiver	482	98
Third-party insurance		
Third-party insurance	17	3
No third-party insurance	470	96
Quarter of initial program enrollment		
Q1 2015	238	48
Q2 2015	249	51

Source: Mathematica analysis of information from awardee's finder file and from Medicare claims and enrollment data as of June 30, 2015.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. Medicaid beneficiaries must be eligible for at least 90 days in the baseline year and on the enrollment date to be included in the sample. Beneficiary characteristics other than CDPS category and risk score were measured in the last month of the baseline year.

CDPS = Chronic Disability Payment System; FFS = fee-for-service; HCBS = ; SSI = Supplemental Security Income; TANF = Temporary Assistance for Needy Families.

Table III.5. Distribution of CDPS categories for Medicaid FFS and managed care beneficiaries enrolled in Montefiore Medical Center's program through June 30, 2015

	All enrollees (N = 487)	
Characteristics	**Number**	**Percentage**
CDPS category[a]		
Beneficiaries in one or more CDPS categories	401	82
Psychiatric	258	52
Cardiovascular	181	37
Pulmonary	137	28
Skeletal and connective	132	27
Gastrointestinal	117	24
Diabetes	103	21
Renal	74	15
Substance abuse	65	13
Central nervous system	60	12
Metabolic	44	9
Genital	40	8
Skin	38	7
Eye	33	6
Infectious disease	33	6
Cancer	28	5
Hematological	22	4
Pregnancy	20	4
Cerebrovascular	7	1
Developmental disability	0	0
Beneficiaries not in a CDPS category	86	17
CDPS risk score[a]		**Value**
Mean		1.86
25th percentile		0.74
Median		1.35
75th percentile		2.38
BHIP target chronic conditions (N = 455)[b]		**Percentage**
Beneficiaries with any of the four target conditions		80
Depression		66
Anxiety		34
Alcohol use disorder		4
ADHD		2

Table III.5 *(continued)*

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of June 30, 2015.

[a]Categories and risk scores are defined by using the CDPS software and are based on the ICD-9 codes on Medicaid claims occurring during each beneficiary's baseline year. The five most common conditions within the sample are reported.

[b]Excludes Medicaid beneficiaries enrolled in a state plan, or who had partial benefits or third-party benefits during the month in which they enrolled in the program. Beneficiaries may have more than one target condition.

CDPS = Chronic Disability Payment System; FFS = fee-for-service.

Consistent with their poor health status and their greater health care needs, both non-dual and dual Medicaid beneficiaries participating in the BHIP had high rates of service use and costs in the 365 days before enrollment. Tables III.6 and III.7 show the baseline cost of care—in total and by major types of services—calculated as the average PBPM Medicaid payment for non-dual and dual beneficiaries, respectively.

Table III.6. Baseline year costs and health care utilization for Medicaid non-dual beneficiaries enrolled in Montefiore Medical Center's program through June 30, 2015

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees[a]	365	327	339	359	365
Average Medicaid costs PBPM[b]					
Total payment	1,194 (144)	1,050 (141)	1,133 (186)	1,268 (211)	1,256 (1202)
Acute inpatient stays	371 (73)	348 (86)	352 (118)	329 (84)	416 (998)
Total ED payment	29 (3)	28 (4)	24 (4)	30 (9)	35 (27)
ED visits that lead to an inpatient stay	3 (1)	3 (1)	3 (2)	1 (<0.5)	3 (14)
ED visits that don't lead to an inpatient stay	27 (3)	26 (4)	21 (3)	28 (9)	32 (14)
Pharmacy	344 (88)	241 (80)	319 (116)	468 (179)	332 (85)
Other[c]	449 (40)	433 (51)	437 (42)	442 (45)	472 (158)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions	441 (79)	488 (99)	355 (106)	444 (119)	432 (1014)
Total ED visits	1,398 (123)	1,361 (166)	1,212 (166)	1,473 (302)	1,506 (1525)
ED visits that lead to an inpatient stay	215 (48)	218 (53)	171 (59)	210 (71)	244 (1013)
ED visits that don't lead to an inpatient stay	1,184 (101)	1,143 (151)	1,040 (136)	1,263 (292)	1,263 (531)

Table III.6 *(continued)*

Source: Mathematica analysis of information from awardee's finder file and Medicaid claims and enrollment data as of June 30, 2015. New York State provided Medicaid data for services through June 30, 2015.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary was enrolled in Medicaid during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Excludes Medicaid beneficiaries enrolled in a state plan, or who had partial benefits or third-party benefits during the month in which they enrolled in the program.

[b]Total Medicaid expenditures for the baseline year or a given quarter exclude capitated payments and were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[c]Other services include practitioner visit services that are not delivered in an office as well as dental, eye care, home health, laboratory, intermediate care facility, nursing home, child care, and clinic services.

ED = emergency department; PBPM = per beneficiary per month.

Table III.7. Baseline year expenditures and health care utilization for Medicaid dual beneficiaries enrolled in Montefiore Medical Center's program through June 30, 2015

		Expenditures and utilization for each quarter in the 12 months before enrollment			
Types of expenditures and utilization measures	12 months before enrollment	Quarter 1 (10 to 12 months before enrollment)	Quarter 1 (10 to 12 months before enrollment)	Quarter 1 (10 to 12 months before enrollment)	Quarter 1 (10 to 12 months before enrollment)
Total number of enrollees[a]	90	87	88	89	90
Average Medicaid expenditures PBPMb					
Total payment	1,304 (186)	1,156 (173)	1,178 (167)	1,156 (166)	1,707 (425)
Acute inpatient stays	135 (92)	29 (11)	46 (16)	56 (17)	401 (364)
Total ED payment	3 (1)	5 (2)	2 (1)	2 (1)	3 (2)
ED visits that lead to an inpatient stay	0 (<0.5)	0 (<0.5)	0 (<0.5)	0 (<0.5)	0 (<0.5)
ED visits that don't lead to an inpatient stay	3 (1)	5 (2)	2 (1)	1 (1)	3 (2)
Pharmacy	79 (32)	76 (32)	79 (30)	69 (32)	93 (40)
Other[c]	1,087 (143)	1,046 (165)	1,051 (162)	1,030 (155)	1,210 (174)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions	560 (131)	373 (125)	692 (218)	684 (195)	492 (165)
Total ED visits	881 (248)	933 (253)	369 (124)	911 (313)	1296 (498)
ED visits that lead to an inpatient stay	114 (46)	140 (103)	92 (64)	182 (89)	45 (45)

Table III.7 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment: Quarter 1 (10 to 12 months before enrollment)	Quarter 1 (10 to 12 months before enrollment)	Quarter 1 (10 to 12 months before enrollment)	Quarter 1 (10 to 12 months before enrollment)
ED visits that don't lead to an inpatient stay	766 (244)	793 (237)	277 (109)	729 (305)	1252 (497)

Source: Mathematica analysis of information from awardee's finder file and Medicaid claims and enrollment data as of June 30, 2015. New York State provided Medicaid data for services through June 30, 2015.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary was enrolled in Medicaid during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Excludes Medicaid beneficiaries enrolled in a state plan, or who had partial benefits or third-party benefits during the month in which they enrolled in the program.

[b]Total Medicaid expenditures for the baseline year or a given quarter exclude capitated payments and were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[c]Other services include practitioner visit services that are not delivered in an office as well as dental, eye care, home health, laboratory, intermediate care facility, nursing home, child care, and clinic services.

ED = emergency department; PBPM = per beneficiary per month.

For non-dual Medicaid beneficiaries, the total average PBPM Medicaid payment during the baseline year was $1,194 (Table III.6). Quarterly estimates ranged from as low as $1,050 in the first quarter to as high as $1,268 in the third quarter, with no discernable pattern over time. The largest drivers of the total cost of care were payments for acute inpatient ($371) services, pharmacy services ($344), and "other" ($449) services. In combination, these services represented 97 percent of the total cost of care.

The lower panel in Table III.6 shows that, for non-dual Medicaid FFS and managed care beneficiaries, the average rate of acute hospital admissions was 441 per 1,000 non-dual Medicaid beneficiaries per year during the baseline year. Quarterly rates ranged from as low as 355 in the second quarter to as high as 488 in the first quarter. The total average rate of ED visits was 1,398 per 1,000 beneficiaries per year. The total annual rate of ED visits that led to an inpatient stay was 215 per 1,000 beneficiaries per year. The rate of ED visits that did not lead to an inpatient stay was 1,184 per 1,000 beneficiaries per year. Quarterly rates ranged from as low as 1,212 in the second quarter to as high as 1,506 in the fourth quarter. There was no discernable pattern in the rates for either hospital admissions or ED visits over time.

For dual Medicaid beneficiaries, the total average PBPM Medicaid payment during the baseline year was $1,304 (Table III.7). Quarterly estimates ranged from as low as $1,156 in the first and third quarters to as high as $1,707 in the fourth quarter, with no discernable pattern over time. The largest driver of the total cost of care was "other" services ($1,087)—including, care in nursing homes and intermediate care facilities, which together represented 83 percent of the total cost of care.

The lower panel in Table III.7 shows that the average rate of acute hospital admissions was 560 per 1,000 dual Medicaid beneficiaries per year during the baseline year. Quarterly rates ranged from as low as 373 in the first quarter to as high as 692 in the second quarter. The total average rate of ED visits was 881 per 1,000 beneficiaries per year. The rate of ED visits that lead to an inpatient stay was 114 per 1,000 beneficiaries per year. The rate of ED visits that did not lead to an inpatient stay was 766 per 1,000 beneficiaries per year. Quarterly rates ranged from as low as 369 in the second quarter to as high as 1,296 in the fourth quarter. There was not a discernable pattern in the rate of hospital admissions or ED visits over time.

Montefiore Medical Center expected to reduce preventable hospitalizations and ED visits among Medicaid beneficiaries compared with what the rates would have been in the absence of the BHIP by offering more comprehensive follow-up and preventive primary care than Medicaid FFS and managed care beneficiaries typically received. If the awardee achieved this goal, then a decline in Medicaid expenditures for BHIP participants in the post-intervention period should become evident and would be reflected in our impact estimates.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Montefiore Medical Center worked with two managed care organizations to develop a monthly bundled payment model to support the implementation of the BHIP. The managed care organizations provide coverage through Medicaid managed care, Medicare Advantage, and commercial plans. As of mid-September 2017, the awardee was in final negotiations with payers. It expected to implement the model at all six participating sites by early 2018.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Montefiore Medical Center engaged with two managed care organizations to develop a monthly bundled payment model to support implementation of the BHIP collaborative care model in 6 of its 22 primary care sites. The managed care organizations provided coverage through Medicaid managed care, Medicare Advantage, and commercial plans. Working with a consultant, the awardee's finance team modeled a variety of approaches based on key program metrics including caseload, engagement rates, type and dosage of treatment, and outcomes. Based on its analyses, Montefiore Medical Center planned to propose monthly bundled payments from payers to sites that would range from $110 to $150 per patient. However, the final payment amount remained undetermined while the awardee continued to negotiate with payers to develop the payment model. Evidence of at least one monthly contact between the patient and a member of the behavioral health team would be required for payment. The target caseload for social workers and consulting psychiatrists would be 120 to 150 patients. Providers could achieve an additional 15 percent bonus for meeting quality and performance targets, including (1) at least 75 percent of patients receive at least one follow-up contact after initial assessment, (2) at least 50 percent of patients show 50 percent improvement on symptom scales by the 10th week, and (3) at least 75 percent of patients who show no clinical improvement by Day 70 receive a consultation or case review.

In addition to monthly bundled payments covering eligible Medicaid, Medicare Advantage, and commercially insured patients, the awardee will also offer program services to Medicare FFS beneficiaries in its Next Generation Accountable Care Organization (ACO). Montefiore Medical Center will assess quality metric performance for this patient population separately, and then pay bonuses to primary care sites through their shared savings agreement.

C. Status of the payment model

Montefiore Medical Center finalized data use agreements with two health plans and worked with a financial consultant to develop and test the model at two sites. The awardee did not engage in developing the model with a third payer as originally planned because of leadership changes at the plan. As of mid-September 2017, the awardee was conducting final negotiations

with payers. It expected to implement the model at all six sites by early 2018. After reaching a final agreement with its payer partners, the awardee anticipated dissemination of a payment model template to engage additional payers to help sustain and scale the BHIP. The awardee continued to work closely with the New York State Office of Mental Health to align the program with the state's best practices for collaborative care and metrics and to receive feedback to inform development of a sustainable payment model.

D. Factors associated with the development of the payment model

The awardee described several factors that facilitated development of a payment model for the BHIP. Montefiore Medical Center has experience with risk-based payment models and long-standing relationships with participating payers, who recognized the importance of integrated care and agreed to participate in the formulation and virtual testing of the model. Further, the awardee's working relationship with these large managed care organizations encompasses multiple populations including Medicaid, Medicare Advantage, and commercially insured patients, which offered the opportunity to coordinate negotiation of payment rates. Montefiore Medical Center held early conversations with payers regarding aims and target outcomes to improve quality care for patients, which program leaders described as helpful in increasing payer buy-in and participation. The awardee also maintained payer involvement throughout each stage of model development, provided regular updates on program progress, and included payers in key stakeholder meetings. Finally, CMMI's investment offset program start-up costs and supported Montefiore Medical Center's model design and testing.

The awardee encountered challenges related to the data required to evaluate cost savings and payment model impacts. Given a late program start date of February 2015, the awardee did not have one full year of claims until halfway through Year 2. The claims data had a lag of three months or more. In addition, the sample size was judged inadequate for detailed analysis of costs. The awardee plans to continue to collect more data to increase the sample size for expanded analysis of program cost savings.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Montefiore Medical Center planned to use its no-cost extension to support the six participating primary care sites in their efforts to sustain the BHIP and to scale the program to other Montefiore primary care sites. The awardee specifically reported that it would use the no-cost extension to continue (1) developing the payment model, (2) engaging site leaders, and (3) testing and improving program technology. During the third program year, awardee leaders reported that primary care leaders at sites not currently participating in the BHIP (referred to as non-BHIP sites) expressed interest in implementing the program. Program leaders expected that the BHIP's alignment with the DSRIP would make both implementing and non-implementing sites more interested in sustaining and scaling the program.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Montefiore Medical Center's sustainment strategy included engaging payers, in addition to leaders and staff at implementing organizations. Alignment with external incentives helped the awardee establish formal contracts with three payers that were interested in sustaining the program and helped garner support from leaders at participating

BHIP sites. Moreover, throughout the cooperative agreement, the awardee encouraged sites to operate independently to prepare them for sustaining the program without the awardee's support after the end of the cooperative agreement. The awardee also worked to reduce staff burden by analyzing program data to improve workflows and by employing the use of telemedicine tools in order to improve the efficiency of program operations.

During the second program year, Montefiore Medical Center had scaled its program by integrating behavioral health staff into more primary care sites beyond those in the cooperative agreement.

C. Implementing the SSR plan: progress and changes

Sustainability. Leaders at implementing sites we interviewed reported that the program would be sustained at their site, although with some differences. An administrator at one site noted that the site may not be able to sustain the program as robustly as it ran during the cooperative agreement, but would make sure that important services were sustained, such as psychiatric consultation and medication management. An administrator at another site believed that the program services would also be sustained, although the site may reform the services over time to better meet the needs of the population—such as providing services via tele-health consultations, expanding the scope of services, or encouraging more collaboration between providers.

CMMI granted Montefiore Medical Center a 12-month, no-cost extension, during which awardee leaders planned to focus on sustaining the program in the six project sites and to expand the program to additional primary care sites. To sustain the program, the awardee wanted to focus on how implementing sites could fund the program with fidelity to the model while using Medicaid reform dollars. The awardee also reported plans to approach related health plans to gauge their interest in participating in the payment model.

Montefiore Medical Center continued making progress on engaging health system leaders and leaders at implementing sites, and planned to do so during the no-cost extension. The awardee focused on demonstrating to primary care leaders the financial value of the program, specifically leveraging the BHIP's alignment with the DSRIP. According to awardee leaders, DSRIP funds can cover some program costs such as patient educator salaries. "As you know, it all comes down to dollars and cents," an awardee leader said about how to best engage leaders.

Montefiore Medical Center also encouraged program sustainment by improving aspects of service delivery. For example, awardee leaders reported calculating an optimal caseload that not only covers the cost of services provided but also "earns a little bit of margin as well." The awardee also reported improving technology and tele-health services, including integrating the patient registry into its EMR, and improving the IVR and smartphone application.

> "Across the board, we've been really trying to teach those best practices and start to integrate different elements that we've found to be particularly helpful with the model into the different MMG sites."
>
> —Awardee leader

Scalability. The awardee continued to approach new sites in its health system to implement the program during the third program year. Through teaching non-implementing sites best practices from the BHIP, awardee leaders reported hearing that the sites found many features of the BHIP

valuable, such as how the five-item screen assesses for a broader range of behavioral health disorders, rather than just depression. Although these sites may not implement the BHIP exactly, awardee leaders believed that the programs would embody the BHIP's core elements of evidence-based screening, measurement-based treatment and follow-up care, and psychiatric consultation.

Replicability. Montefiore Medical Center did not report plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

The BHIP's alignment with the direction of health care reimbursements in New York State has greatly facilitated interest from implementing and non-implementing sites at Montefiore Medical Center. The DSRIP incentivizes programs that integrate behavioral health, and specifically offers reimbursements for telephonic care management as well as efforts to push value-based payments and behavioral health collaboration, according to the awardee.

Montefiore Medical Center also reported that the no-cost extension would facilitate sustainment by allowing the awardee to better develop ways to support program sustainment at the implementing sites. For example, the awardee will continue developing the alternative payment model; engaging primary care leaders at implementing sites; and testing the recent improvements to its technology, especially the patient registry.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS requests a follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants Montefiore Medical Center's BHIP. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicaid beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: National Association of Children's Hospitals and Related Institutions

April 18, 2018

Joe Zickafoose (Mathematica Policy Research)
Victoria Peebles (Mathematica Policy Research)
Anna Christensen (Mathematica Policy Research)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries. The National Association of Children's Hospitals and Related Institutions received a no-cost extension (NCE) to continue practice transformation and learning collaborative activities through November 2017, complete data analyses through May 2018, and hold a final meeting with sites in June 2018.

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

The National Association of Children's Hospitals and Related Institutions (NACHRI) used funding from HCIA R2 to implement the Coordinating All Resources Effectively (CARE) program to improve the often disjointed and costly state of care for children with the most complex medical conditions, who are referred to as children with medical complexity (CMC) (Table I.1). The program innovated on earlier approaches to care for this population by engaging both hospital-based outpatient practices for CMC and primary care practices across the country in a learning collaborative to implement a framework for care that was tailored to local needs. With the goal of better care for children and their families, the awardee involved 10 children's hospitals in the learning collaborative to implement and enhance care management, care coordination, family engagement, and practice transformation. The awardee and sites used continuous quality improvement methods to monitor and refine implementation. Each hospital worked with one to six primary care practices and its own hospital-based program for CMC, with the exception of one site without a hospital-based program.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The National Association of Children's Hospitals and Related Institutions (NACHRI) sought to achieve three primary goals: (1) improve the experience of care for CMC and their caregivers through the duration of the program, (2) reduce family stress related to health care by 10 percent, and (3) reduce overall medical expenditures by 6.8 percent.
Major innovations	• Conducted a learning collaborative, including hospital-based practices and primary care practices across multiple states, to implement a package of interventions to improve care for CMC with tailoring to meet local needs • Used claims data from multiple state Medicaid payers for participating hospital sites to identify participants, evaluate health care use, and negotiate new alternative payment models
Program components	• **Care management and care coordination.** Provided to all participants and their caregivers through collaboration between hospital-based staff, hospital- and practice-based care coordinators, and staff in collaborating primary care practices • **Practice-based quality improvement and transformation.** Provided support for primary care and hospital-based complex care practices to transform care processes for CMC, consistent with the principles of the medical home • **Education and training.** Conducted a learning collaborative based on The Breakthrough Series from the Institute for Healthcare Improvement
Focus population	CMC were defined as children classified into the 3M™ CRG software categories 5b, 6, 7, 8, or 9 by using billing or claims data. These categories encompass children with lifelong chronic conditions, complex chronic conditions, and malignancies.
Theory of change/ theory of action	NACHRI hypothesized that better care management and coordination, heightened family engagement, and practice-based quality improvement and transformation will lead to better care experiences, reduced family stress, and lower costs of providing health care to CMC.
Payment model	Unique to each implementing site, including per capita care management payments, shared savings, and fee-for-service
Award amount	$23,198,916
Effective launch date	5/1/2015
Program setting	Hospitals and primary care practices
Market area	Urban, suburban
Market location	California, Colorado, District of Columbia, Florida, Missouri, Ohio, Pennsylvania, Texas

Table I.1 *(continued)*

Program characteristic	Description
Target outcomes	• **Better care.** Improve patient and caregiver experience. • **Healthier people.** Reduce family stress related to care by 10%. • **Smarter spending.** Reduce medical expenditures by 6.8%.

CARE = Coordinating All Resources Effectively; CMC = children with medical complexity; CRG = Clinical Risk Group.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was effective in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, the awardee enrolled 8,111 participants—100 percent of its enrollment target—by the end of the initial cooperative agreement. Second, the awardee's sites exceeded its goals for engaging practices by enrolling a total of 44 primary care practices and 9 hospital-based complex care practices. Third, the awardee reported that most practices were implementing the program's services according to the awardee's metrics. Fourth, site leaders and staff reported that participants' families were engaging with the program. Finally, site leaders and staff perceived that the program had a positive effect on the delivery of care.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of the National Association of Children's Hospitals' CARE program, the analysis is still in progress and not included in this report.

Payment model. At the time of final interviews, 3 of the 10 awardee sites had contractual agreements with the state Medicaid agency or a Medicaid managed care organization (MCO). After our interviews, the awardee reported that one additional site had reached an agreement with a Medicaid MCO, two sites were in active negotiations with payers, and the remaining four sites continued to analyze data and develop payment model proposals. The four payment model agreements reached during the cooperative agreement included two with per-beneficiary-per-month (PBPM) care coordination payments, one with shared savings in addition to fee-for-service, and one with an upfront infrastructure payment for a care coordination team with the potential for shared risk.

Sustainability plans. The awardee was actively working with its 10 implementing sites to develop sustainability plans and payment models to support the CARE program. At different phases of the implementation, the sites began a variety of activities to promote sustainability—including, assessing which parts of the program were most effective, exploring how to staff and embed program components in existing functions, assessing support from the site's parent organizations, and identifying and securing resources necessary for long-term sustainability. Several of the sites were in the process of scaling the program to other parts of their systems. The awardee and the sites also worked to disseminate information about the program; other children's hospitals had expressed interest in the program, but none had specifically replicated it. The awardee planned to use part of its 10-month NCE to help further sites' sustainability plans and the activities promoting scaling and replication.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of two data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. For this awardee, we interviewed the leaders of the program and leaders and staff from at least three of the program sites in each round of interviews. Because we were unable to obtain valid sample frames—either through the awardee or from sites of clinicians and non-clinician staff working on the award—we were unable to field the surveys of clinicians and non-clinician staff with this awardee.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a] Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

Through the CARE program, the awardee focused on children with the most complex chronic conditions, which typically involved multiple organ systems and required the services of many medical specialists. The awardee sought to enroll about 8,064 Medicaid-enrolled CMC in the program across the 10 participating hospital sites.

For the purposes of the program, CMC were identified based on categories from the 3M™ Clinical Risk Group (CRG) algorithm for billing or claims data. Specifically, the awardee focused on children who were classified in the highest categories of the algorithm, which encompass children with lifelong chronic conditions, complex chronic conditions, and malignancies (CRG groups 5b, 6, 7, 8, or 9). However, the awardee and sites were not able to use Medicaid data to identify potential participants in the first year of the program because of the

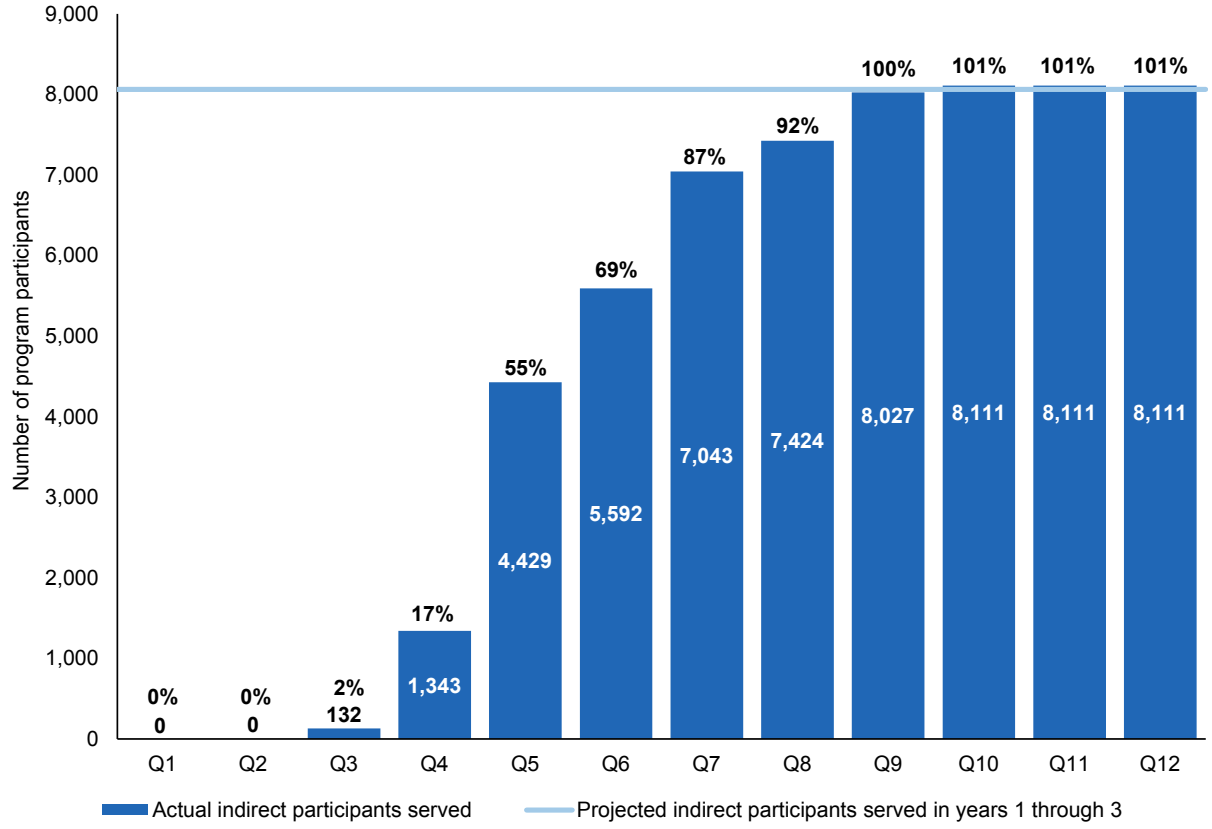
time it took to negotiate with state Medicaid agencies and MCOs for access to data and because there were lags in the time periods covered by the data once access was obtained.

Participating hospitals identified potential participants and enrolled them using data from any of three sources: (1) existing hospital practices for CMC, (2) analyses of administrative and billing data from the children's hospitals, and (3) referrals from providers. CMC who receive care at participating complex care and primary care practices are "passively enrolled" into the program. Children who were already receiving care coordination were also passively enrolled into the quality improvement efforts for care coordination. For children who were not already receiving care coordination, parents had to agree for them to participate in care coordination, but there was no formal consent process. Once the awardee obtained Medicaid data for a site, its team used the data to describe the CRG categories of participants after the fact of their enrollment.

b. Evidence of enrollment effectiveness

National Association of Children's Hospitals was effective in meeting its enrollment target by the end of the 12th program quarter. Overall, the awardee reported that it enrolled 8,111 participants from May 2015 (when it launched its program) through August 2017, which represents just over 100 percent of its 8,064 three-year projected participants (Figure II.1). At the end of the first year of the cooperative agreement, the awardee had enrolled 17 percent of its target, but that had increased to 92 percent of its target by the end of the second year. At the beginning of the cooperative agreement, the awardee's cumulative target was 9,041, which was later revised to 8,064 based on sites' reassessments of the number of potential participants enrolled in Medicaid.

Figure II.1. Projected versus actual cumulative indirect participants served through Year 3 as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Participants are referred to as indirect for NACHRI because the awardee sites were the direct service providers, not the awardee. The awardee lowered its indirect participant enrollment projection from 9,041 at the beginning of Year 1 to 8,032 at the beginning of Year 2. It then increased the projection slightly to 8,064 at the beginning of Year 3.

c. Barriers and facilitators associated with enrollment effectiveness

The National Association of Children's Hospitals and Related Institutions' progress in meeting its three-year enrollment goals was influenced by several factors. The awardee was able to address its challenges and meet its enrollment goals by the end of the second year of the cooperative agreement.

The awardee and site leaders reported three primary barriers to enrollment: (1) the Institutional Review Board (IRB) review and consent processes at the sites, (2) the multiple processes needed to identify eligible children, and (3) the time needed to recruit and contract with independent primary care practices. First, as described in the second annual evaluation

report,[3] the IRB and consent processes were particularly arduous in the 18 months of the award, and contributed to delays in enrollment. The awardee and sites eventually addressed this problem by deciding to consent only the one-third of participants who would receive a program survey and passively enrolling other participants if they were receiving care at a participating practice site.

Second, several site leaders talked about the problems involved in identifying eligible children. The awardee and many sites initially planned to use Medicaid claims data to identify eligible children who received care at participating practices, but the process required to obtain claims data took the majority of the first year for most sites. As a result, most sites used a combination of internal billing data, manual chart reviews, and Medicaid data to identify eligible children. Although the sites were able to develop effective processes, these processes were more time-consuming than anticipated.

Third, for sites that did not have affiliated primary care practices or that chose to work with independent practices, the process of engaging independent practices was more time-consuming than anticipated. The sites were unable to enroll children from the practices until they had a contractual agreement, and in many cases, negotiating a contract took much of the first program year. However, awardee and site leaders reported that they had reached agreements with all participating practices as of early 2016.

The awardee and site leaders also said there were two primary facilitators to enrollment: (1) placing renewed emphasis on enrollment during the second program year, and (2) leveraging existing relationships with primary care practices and health plans. After challenges with enrollment in the first year of the cooperative agreement, awardee leaders began giving a monthly feedback dashboard to executive sponsors and teams for each site to share with chief executives and other stakeholders in order to compare enrollment progress with the goals at each site. Carryover funds from the first year were also allocated to help support additional time spent by staff at the sites on enrollment. Program and site leaders credited this renewed emphasis for the significant gains in enrollment during the second year of the cooperative agreement.

For several sites, enrollment efforts were facilitated by leveraging relationships with owned or affiliated primary care practices. For example, two of the sites that made the fastest progress on enrollment were working exclusively with primary care practices owned by their health systems, which allowed for faster enrollment and engagement with the practice. In contrast, sites working with independent primary care practices typically had slower enrollment due to delays in contracting with these sites.

[3] Zickafoose, Joe, Victoria Peebles, Eric Lammers, and Anna Christensen. "HCIA Round Two Evaluation: National Association of Children's Hospitals and Related Institutions." In Gilman, Boyd, Grace Anglin, Karen Bogen, Arnold Chen, Leslie Conwell, Stacy Dale, Laurie Felland, Kristin Geonnotti, Rachel Kogan, Nancy McCall, Dana Peterson, Rumin Sarwar, Bob Schmitz, Ellen Singer, Joe Zickafoose, Theresa Feeley-Summerl, Cyrus Jadun, and Randall Brown. "Evaluation of the Round Two Health Care Innovation Awards (HCIA R2): Second Annual Report." Report to the Centers for Medicare and Medicaid Services. Cambridge, MA: Mathematica Policy Research, August 2017.

2. Delivery of program services

a. Description of and changes to service delivery model

The awardee's primary intervention strategies for patient care were care management and care coordination provided to all participants. Depending on the site, these were delivered by two groups: (1) hospital-based staff, including care coordinators and social workers, and (2) staff in collaborating practices—or through collaboration between the two groups. The CARE program included five care coordination and management processes, or "change concepts," for implementation and quality improvement at participating hospitals and primary care practices: (1) a patient registry, (2) shared identification between families and providers of members of the child's care team ("dynamic care team"), (3) access plans, (4) care plans, and (5) transitions of care. These change concepts were initially developed after a review of the literature and consultation with the program's expert faculty. The change concepts were refined in collaboration with the sites. For each change concept, program staff developed guiding principles, common core elements of change, suggested actions, and quality measures. Program staff supported participating hospitals and primary care practices in implementing the five change concepts through the learning collaborative, and the sites hired practice transformation facilitators to support local implementation and tailoring of the concepts. The program also encouraged sites to engage caregivers in more formal advisory roles and get their input on the implementation and refinement of the program.

The awardee and sites tailored aspects of the change concepts to address site specific needs, but the core concepts stayed the same throughout —with the exception of the concept on transitions in care. This concept was originally thought of as the transition of care for patients with less complex or more stable conditions from the hospital-based practices to primary care practices. However, early in the award, the sites noted that the participating primary care practices were already caring for significant numbers of CMC, so the concept was removed from the collaborative effort.

In interviews, site leaders and staff reported that they had been using many of the care coordination and management processes before the award, especially in the hospital-based practices, but the CARE program encouraged them to be more systematic in what they were already doing and to adopt new processes to strengthen current practice.

By engaging with hospital-based practices for CMC and primary care practices, the awardee sought to increase the frequency and improve the quality of care management and coordination processes. In improving these processes, the awardee intended for the sites to work with families to identify and address unmet needs, give clearer plans for communication and seeking care in the event of acute needs, offer more support for self-management, and improve communication between providers. As a result, families would have better experiences with care and lower stress levels related to care, and improvements in the child's care would result in lower health care spending by decreasing the incidence of avoidable utilization, such as emergency department visits and hospitalizations.

b. Evidence of service delivery effectiveness

After initial delays in enrollment, National Association of Children's Hospitals was effective overall in implementing its service delivery model. As we describe below, the awardee reported

steady increases in sites' implementation of the change concepts for its model, in expected levels of staffing at the sites, and in accomplishment of goals for practice enrollment and engagement. Site leaders and staff reported that participants' families seemed well engaged with the program. The awardee, site leaders, and site staff described how earlier challenges with the program had largely been overcome, and ongoing challenges with maintaining clinical champions and engaging families at an appropriate level were being managed.

Delivery of intervention services. In interviews, program and site leaders reported that, due to delays in enrolling participants, the program was not delivering services to the full extent intended until its second year. They described the program as "hitting its stride" sometime between summer and fall of 2016, meaning that enrollment was at a high enough level, and the interventions had matured satisfactorily.

Beginning in May 2016, the awardee began to report back to sites about their performance on measures of implementation related to three of the change concepts: identification of members of the child's care team ("dynamic care team"), access plans, and care plans. The awardee also reported on the presence of a patient registry, but stopped reporting this when all practices reached this expectation. These metrics were included in the monthly executive dashboard sent to sites that also included enrollment. Both hospital-based complex care practices and primary care practices showed steady improvement on the three measures from May 2016 through June 2017, although performance on specific measures varied across hospital sites and practice types.

Beyond the awardee's standardized reporting of specific change concepts, it was difficult to compare implementation at different hospital sites because the awardee allowed sites to determine the details of many aspects of the change concepts. For example, the program did not establish minimum standards for follow-up with participants once care coordination began. At some sites, care coordinators only followed up with caregivers if the child had an acute event such as a hospitalization, or if there was a specific request from the caregiver; however, at least one site established standardized follow-up based on the complexity of needs, with the follow-up period ranging from weekly to monthly.

Recruitment and engagement of providers. As of May 2017 (near the end of the cooperative agreement), the awardee had 42 primary care practices and 9 hospital-based complex care practices that were participating and submitting data monthly, which met the awardee's goal of 1 to 8 practices per hospital site and a total of 40 practices. The number of practices that were engaged in the program had increased from 40 in early 2016 for primary care practices, and had remained steady for hospital-based complex care practices. One hospital site did not have a hospital-based outpatient program for CMC, and was only implementing the program through primary care practices. Most sites worked primarily with owned or affiliated primary care practices, but some were able to include independent practices. Six primary care practices across the 10 sites dropped out over the course of the cooperative agreement, and are not included in the practice counts from May 2017. The awardee leaders said these practices did not feel they had the staff resources to meet the program's expectations.

Staffing and training. Sites hired staff for the program. Awardee and site leaders reported that they were able to meet their staffing needs, but did not report specific staffing goals or

numbers for the sites. They did say staff turnover was a challenge, especially in the care coordinator role. They noted that staff turnover for the program was not any greater than what they experienced elsewhere in their organizations, but the impact was amplified in practices that typically have a small number of staff. However, the awardee and site leaders thought the affected teams were able to adapt, continue delivering services, and eventually fill the empty positions.

Engagement of program participants. The program engaged participants' families in two ways: (1) shared development access and care plans, including identification of care teams, and (2) family advisory groups. First, engaging families was a central part of the direct services in the model, and it was intended to be accomplished in particular by collaborative creation of care plans by care coordinators and families. All interviewees said this process helped engage families they were able to reach, and the number of enrolled children with completed access and care plans steadily grew (Figure II.2).

Second, the awardee encouraged sites to involve families as program advisors. The awardee reported that most practice sites were able to incorporate individual family advisors or advisory groups into their programs. The site leaders and staff described how the advisors made important contributions in helping design and refine specific processes and documents, such as care plans.

c. Barriers and facilitators associated with service delivery effectiveness

In the first two years, the awardee's service delivery was delayed by the initially slow pace of enrollment, the complexity of obtaining IRB approval from multiple sites, and the difficulty of obtaining Medicaid data and engaging community practices, but by the third year of the cooperative agreement, the awardee had overcome these challenges. In the third year, awardee and site leaders described two major obstacles they had dealt with: (1) engaging and maintaining clinical and administrative champions, and (2) balancing engaging families with the program with limiting their reliance on the program for less complex issues. First, awardee and site leaders noted the ongoing importance of both clinical and administrative champions. Site leaders said that when practice champions left their jobs or clinical champions dedicated too little time to the project, engagement with practices suffered, but they also talked about how their recognition of the importance of champions led them to recruit new champions and to get more time for the clinical leaders to dedicate to the program. Awardee and site leaders also said executive sponsors and other administrative champions were important throughout in prioritizing needed changes in electronic medical records (EMRs), authorizing staff positions, and supporting other program needs.

Second, site leaders did note that some families were difficult to reach, and some declined to enter the program if their child had less complex conditions that still met program enrollment criteria (for example, obesity and asthma). Still other families seemed overwhelmed by the idea of having one more person involved in their child's care. Sites addressed these challenges in a variety of ways, including completing and mailing care plans to families they could not reach and letting families know they could join the program later if they changed their minds. In contrast with the families who were tough to engage, other families (according to some site leaders and staff) began to contact the program for issues the staff thought they should be able to manage on their own, such as contacting a medical supply company to find out a delivery date or

contacting an interpreter. The staff responded by telling the family how to do the task on their own and by including instructions in the care plan about the best contact for a given issue.

C. Assessment of perceived program effects on the delivery of care and outcomes

In interviews, awardee leaders and site leaders and staff talked about how the program had affected care in different ways. Three of the most commonly reported effects were: (1) greater access to advice and more support on self-management for families with concerns, (2) embedding the tools and processes from the program in routine care, and (3) data from hospital sites that suggested there were fewer hospitalizations and shorter stays, but did not reveal clear changes in the number of ED visits. First, a number of interviewees said the program had given families better access to clinical advice and had emphasized teaching families what to do in response to specific concerns. They said the practice sites have made it easier to get in direct contact with a member of the care team, most commonly the care coordinator. They also talked about how care plans and their interactions could support families in managing concerns on their own.

> "I think that we're doing a better job of educating them [families] in terms of their disease processes and their health care team and who's responsible for which… [for example,] Is this something you should call your regular doctor for? Do you think this is something … to call your neurologist [for] or [go] to the ER?"
>
> —Program staff member

Second, several interviewees described how the program has allowed some practices to serve as "learning labs" to develop processes and tools that can be spread to other practices and hospital programs. They told us how tools built for the program, such as care plans and lists of the dynamic care teams, began to be used routinely by participating practices. This was especially true of tools embedded in an EMR. These tools became visible and usable for those in other settings—such as the emergency department and specialty practices—allowing them to quickly grasp the full spectrum of the child's conditions and plans of care.

Third, some sites reported that their review of hospital data comparing trends before and after the program started suggested some decreases in participants' hospitalization rates and lengths of hospital stay. The site leaders noted they have seen less change in ED visit rates, and some attributed this to the complexity of the children's conditions and the need to respond to many of their acute needs in the ED setting. The site leaders had not tracked other potential drivers of care spending, including pharmaceuticals and medical supplies, and noted that the program could increase spending in these categories because there would be better identification and management of participants' unmet needs.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. The results of our implementation evaluation have several implications for any future impact evaluations. First, the level of complexity in enrolled children's needs is likely to vary by site. Although the program had consistent eligibility criteria for participants based on claims data, implementation of these criteria at the site level could vary with the data available to the site and with Medicaid enrollment criteria. Most sites did not have access to Medicaid claims

early in the program, and thus needed to use a combination of existing enrollment in complex care programs, hospital billing data, and, in some cases, manual review of patient medical records. Also, some sites only enrolled children in Medicaid managed care, and some of the children who had the most complex conditions and may have been eligible for the program were excluded from managed care.

Second, some sites reported possible trends toward fewer hospitalizations and shorter lengths of stay in the hospital. Trends in ED visits were unclear, and they had no information on use of other health services. As a result, awardee and site leaders were less sure about the likelihood that the program would be associated with lower overall spending on participants' health care.

Third, although the program consisted of a standardized set of change concepts, individual sites had substantial freedom in implementing the details of these concepts, and they had varying levels of success meeting some of the awardee's standards. The resulting variation between sites on many details of the program could affect the impacts of the program, such as how often care plans incorporated all the criteria expected by the awardee and how often care coordinators followed up with families.

Fourth, the awardee and site leaders consistently asserted that most of the first year was spent developing enrollment processes, engaging practices, and refining the service delivery model, and they expected it would take 6 to 12 months for the program to affect participant outcomes. As a result, they did not expect to begin to see program effects until about two-thirds of the way into the second program year, the summer of 2016.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of the National Association of Children's Hospitals' CARE program.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: National Association of Children's Hospitals

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018	8,116[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	4,660
Likelihood of all-cause hospitalizations	580
MDE sample size requirement to detect 20% effect	
Total expenditures	1,165
Likelihood of all-cause hospitalizations	145
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group; patient self-selection high
Intervention implemented that differs from baseline period	Fully implemented intervention with differences in services from baseline that varied by site
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	Too early to determine due to delay in Medicaid data
Do claims identify the primary expected effects?	Some effects observed in claims data but important effects likely missing
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	None
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We will be able to conduct a rigorous impact evaluation of the CARE program on health care utilization. We will match children in the treatment group from each hospital to a group of comparison children in the same state who have similar demographic and diagnostic characteristics and similar health care utilization patterns in the year prior to enrollment. Table III.1 shows our final projected sample size, which reflects the number of treatment group children at the time enrollment when it ended in August of 2017. The sample should be large enough to detect plausible effects of the program, though it is possible that lags in Medicaid data in some states could prevent us from doing an analysis of beneficiaries from some of the participating hospitals.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

At the time of final interviews, three of the 10 awardee sites had contractual agreements with the state Medicaid agency or a Medicaid MCO. After our interviews, the awardee reported that one additional site had reached an agreement with a Medicaid MCO, two sites were in active negotiations with payers, and the other four sites continued to analyze data and develop payment model proposals. The four payment model agreements reached during the cooperative agreement included two with PBPM care coordination payments, one with shared savings in addition to fee-for-service, and one with an upfront infrastructure payment for a care coordination team, with the potential for shared risk.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

The awardee supported each of the 10 sites (in eight different states) in developing its own proposals. The program was focused on children enrolled in Medicaid, and as a result, each site separately negotiated with its state Medicaid agency or with Medicaid MCOs operating in its region. We focused on the three sites participating in the CARE program that reached contractual agreements before July 2017 (Table IV.1).

After leaders at Children's Mercy Hospital and Clinics in Kansas City, Missouri, engaged with the state Medicaid agency, the agency agreed to adjust the criteria to participate in the state's existing Medicaid health homes program to include additional diagnoses relevant to children, such as developmental disabilities, obesity, and asthma. The health homes program provides a PBPM care coordination fee for individuals with two or more specific chronic conditions or with the stand-alone diagnoses of pediatric asthma, pediatric obesity, or diabetes (Table IV.1).[4] Only about one-fourth of the children enrolled in the CARE program at this site qualified for services under this payment model because of its more restrictive list of diagnoses for eligibility. Site leaders expected the PBPM amount to cover a substantial amount of the program costs but had not done a full cost analysis, because the site's executives were committed to integrating the program's services into routine care regardless of the results of such an analysis.

St. Joseph's Children's Hospital entered into a new agreement for a nominal PBPM care coordination fee with Florida's Children's Medical Services, an existing Medicaid managed care

[4] The Affordable Care Act created an optional Medicaid benefit for states to use health homes to coordinate care for people with chronic conditions. More information about the Missouri Medicaid health homes program is available at https://dss.mo.gov/mhd/cs/health-homes.

program for CMC in Florida. The agreement covers the same population and services as those implemented during the CARE program, and is viewed as a starting point for future negotiations.

Cook Children's Health Care System in Fort Worth, Texas, reached an agreement between its affiliated hospital, provider network, and Medicaid MCO for a shared savings payment model in addition to existing fee-for-service payments. In the model, savings calculated from a baseline period before the intervention would be shared equally between the three organizations. The model has the same focus population and services as the CARE program.

Table IV.1. Payment model characteristics for three implementation sites reaching contractual agreements during the National Association of Children's Hospitals program

Characteristics	Children's Mercy Hospital and Clinics (Kansas City, MO)	St. Joseph's Children's Hospital (Tampa, FL)	Cook Children's Health Care System (Fort Worth, TX)
Payment model	PBPM care coordination fee	PBPM care coordination fee	Shared savings
Payer	State Medicaid agency	Statewide Medicaid MCO for children with medical complexity	Medicaid managed care plan
New/existing payment model	Existing Medicaid health homes program; program newly applied to children	New model within existing contractual relationship with MCO	New model within existing corporate affiliation between the hospital, provider network, and MCO
Differences from program's focus population	Had at least one ED visit or hospitalization in the prior 12 months and had two or more target conditions (asthma, developmental disabilities, diabetes, heart disease, obesity, tobacco use, anxiety, or depression) or had one stand-alone condition (diabetes, pediatric asthma, or obesity)	No differences	No differences
Differences from program services	Existing program services plus more frequent contacts (monthly) and involvement of a behavioral health consultant	No differences	No differences
Payment amount	$63.72 PBPM	$2 PBPM	Shared savings split evenly between the hospital, provider network, and MCO
Adjustments based on patient characteristics	No	No	No
Adjustments based on quality or outcome measures	No	No	No

The awardee used part of its award funding to provide support for actuarial consulting to each site that used Medicaid claims data available to it to develop initial population summaries of spending, to work with the site to select a payment model and design, and to develop an actuarial model of program reimbursement and expenses. The awardee also facilitated peer learning among the sites' leaders about approaches to developing payment models for the CMC population.

C. Status of the payment model

All three sites signed contractual agreements for their payment models with a state Medicaid agency or Medicaid MCO during the second or third years of the cooperative agreement. Children's Mercy planned to discuss options with the Missouri Medicaid agency for expanding the number of children from the CARE program who qualified for the health homes program, and to explore opportunities for an accountable care organization-like arrangement for Missouri Medicaid enrollees based on its experiences with a shared savings arrangement with Kansas Medicaid and value-based models with private payers. St. Joseph's planned to use the agreement for a nominal care coordination fee as a starting point for ongoing negotiations with the Medicaid MCO.

Cook Children's reported that it had not paid any shared savings after more than nine months of its model. Its leaders said there were two reasons for this: (1) initially, Cook Children's enrolled children whose conditions were less complex than expected, because more complex children who were enrolled in Medicaid and Supplemental Security Income (SSI) were excluded from Medicaid managed care until late 2016, and (2) children in the affiliated health plan had lower than national rates for hospitalization and hospital length of stay before the CARE program began.

The site leaders described how children enrolled in Medicaid and SSI in Texas were often those with the most complex conditions, and were thus excluded from managed care until late 2016. As a result, the site leaders believed that participants at the site had medical conditions that were less complex than anticipated, and consequently there were fewer opportunities to reduce their health care spending. Site leaders also noted that before the program started, CMC in Cook Children's health plan already had fewer hospitalizations and hospital lengths of stay compared to national averages, which they felt limited opportunities to reduce health care spending. In late 2016, Texas began to enroll children who qualified for Medicaid based on a disability into managed care, and Cook Children's had planned to apply the same service delivery and payment model to these children. However, program leaders at the site described the state's implementation of this as a "hard override" for their plans because the state reduced the capitation rate to health plans by 8 percent compared to historical spending for this population and required health plans to maintain all existing levels of service for these children. Cook Children's leaders were unsure if they would be able to realize savings under those restrictions.

After the time of our interviews, the awardee reported that one additional site (Children's Hospital of Philadelphia) had reached an agreement with a Medicaid MCO for an upfront infrastructure payment to fund a care coordination team with the potential for shared risk. Two additional sites were in active negotiations with payers, and the remaining four sites continued to analyze data and develop payment model proposals.

D. Factors associated with the development of the payment model

In awardee documents and interviews, leaders at the awardee and the sites identified three common facilitators to developing payment models during the cooperative agreement: (1) having access to a large set of Medicaid claims data for actuarial analysis, (2) building on existing payment programs and relationships with payers, and (3) sharing lessons across the program sites.

First, interviewees described how the work to obtain, standardize, and analyze claims data from all implementing sites and the actuarial contractor supported their efforts to develop payment models. Second, all three sites that had implemented payment models built on existing payment models or relationships with payers. Children's Mercy integrated with an existing payment program, St. Joseph's built on an existing relationship with the state's managed care program focused on children with complex needs, and Cook Children's leveraged its unique affiliation with an MCO. Third, awardee and site leaders extolled the value of the awardee's learning collaborative, which included specific sessions for site leaders to discuss approaches to developing payment models and to share lessons learned.

Awardee and site leaders also reported three major challenges to developing the payment model: (1) a relatively small focus population with a wide range of conditions and spending, (2) the difficulty of obtaining claims data that allowed for meaningful comparisons to populations not enrolled in the CARE program, and (3) the time needed to develop a new payment model. First, interviewees said it was challenging to create actuarially sound value-based payment models for the CARE program's focus population because of its highly variable needs and spending. The awardee reported that this has also resulted in limited interest from some payers. For now, sites have primarily addressed this by working on models that do not include shared savings or adjustments based on spending. The one exception was Cook Children's, and it did not pay out any shared savings over a one-year experience with its model.

Second, although the awardee and site leaders were able to negotiate access to Medicaid claims for all sites, most sites only had access to claims for children who received care at their hospital and, in some cases, for only one MCO. As a result, the awardee and sites could not compare spending and utilization for children enrolled in the program to spending and utilization for children receiving care elsewhere.

Third, some sites were interested in pursuing value-based payment models, but said the delays in obtaining Medicaid data, the existing Medicaid data lags, and the three-year length of the cooperative agreement did not give them enough time to develop a payment model that involved financial risk. Several sites expressed interest in continuing to explore value-based models after the cooperative agreement, possibly joining an existing accountable care organization or starting their own.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The National Association of Children's Hospitals and Related Institutions was actively working with its 10 implementing sites to develop sustainability plans and payment models to support the CARE program. At different phases in developing and implementing the program, the sites started a variety of activities to promote sustainability. These included assessing support from the site's organization, identifying and securing necessary resources, and embedding program components in existing functions. Several of the sites were in the process of scaling the program to other parts of their systems. The awardee and the sites also worked to disseminate information about the program; other children's hospitals had expressed interest in the program, but none had specifically replicated it. The awardee planned to use part of its 10-month NCE to help further sites' sustainability plans and their activities to promote scaling and replication.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, the awardee's implementing sites began using strategies to sustain portions of the CARE program. Program sites demonstrated progress in sustaining care management and coordination services through various activities, including: integrating program services into

existing hospital programs; modifying EMRs to permanently support program services; and securing internal funding for the program—in some cases building on support gained through developing a payment model or getting reimbursement for achieving certification as a patient-centered medical home. However, the sites had either no plans or only nebulous plans for sustaining the program's practice transformation and provider training components. The sites had not scaled the program, nor were there known examples of other organizations replicating the program.

C. Implementing the SSR plan: progress and changes

Sustainability. In Year 3, the sites were assessing the feasibility of embedding the program processes into their existing practice after the award period. As an organizing entity, the awardee's role in sustainability was expected to be small compared to what each site was planning on its own.

Implementing payment models was a large part of each site's sustainability effort, but the 10 sites were each pursuing separate payment models. For longer-term sustainability, as discussed in Chapter IV on the payment model, the participating sites were at different stages of securing contractual agreements with Medicaid agencies and managed care organizations. At one end of the spectrum, four sites had agreements in place, and at the other end, four were still analyzing data and developing proposals.

The awardee asked each site to develop a sustainability plan. The first step involved identifying which change concepts were most essential to sustain. The second step involved reviewing the results of a survey the awardee administered to several staff at each site to get their perspectives on the critical success factors for sustainability, the importance of these factors at their care site, the effectiveness of different components of the program, and their staff members' perceived ability to continue to improve care. The third step was for each site to develop a structured sustainability plan by the end of August 2017.

In addition to covering services through November 2017 and data analyses through June 2018, the awardee planned to use its 10-month NCE to further its SSR activities. These included reviewing the sustainability plans with the sites and providing expert advice on the plans from the CARE program's faculty, as well as developing online micro-learning modules to support scaling and replication.

The awardee and individual hospital sites were working on several ways to help sustain the CARE program. Sustainment strategies encompassed several areas:

- **Defining and setting clear roles and workflows for staff**, particularly for care coordination and care management, functions most sites planned to continue. One site identified the need to improve communication between the main stakeholders—including the PCP sites, clinics, EDs, and medical directors—as a way to understand everyone's roles and ensure seamless patient handoffs in the future. In contrast, another site did not plan to continue the warm handoffs between inpatient and outpatient services and with the community because it found them too difficult to do in real time.

- **Determining the ideal staffing ratio for the program**, based on guidance from the awardee on taking patients' medical and social complexity into account and on any evidence that the intervention reduced the need for intensive coordination and management over time. One site was risk-stratifying the patient population and assigning staffing resources to match the need.
- **Continuing to collect and analyze data** to identify where the program was most successful, and focusing efforts there. One site planned to keep administering the Medical Home Index Survey regularly, because it found the results useful for identifying where the it was excelling and lagging in implementing the change concepts. In that same spirit, another site was employing its experience outside the scope of the award in creating consistent asthma plans across clinics to do the same for the program's emergency access plans.
- **Incorporating program functions into systemwide initiatives and infrastructure.** This included embedding care coordinators into both PCP clinics and complex care clinics. Also, some sites planned to have the care coordinators administer the patient surveys, potentially in a shortened fashion to focus on the highest priority elements (for example, ensuring the patient has a care plan). Several sites planned to embed care coordination tasks and tracking in their EMRs. As a respondent said, "Once it exists in there, then that means it's not going anywhere and it's something that we can [continue to] use."
- **Expanding the role of the program team's members** to help them treat patients more proactively and holistically, including meeting the needs of the patients' families. For example, one site focused on staff working at the top of their license and distributing work most effectively among the members, and identifying areas where it needed to recruit additional team members.

Scalability. Several of the sites were in the process of scaling the program to their whole systems or to other parts of their systems. For example, one site had standardized care management processes within its affiliated Medicaid and CHIP health plan by adopting a care management team concept that it referred to as "pods." In the third year of the cooperative agreement, the health plan had already expanded from five pods under the award to 20 pods enabling it to work with broader populations of children. Other sites were also standardizing care management and coordination processes across all their affiliated primary care and outpatient complex care practices, finding that it was easier for staff to complete care processes when those processes were part of the standard of care for everyone.

Replicability. The CARE program had not been replicated. However, other hospitals reached out to the awardee and its sites to learn more about the program and to discover which results might be transferable if they were to replicate the program. The awardee also proactively worked to disseminate information about the program by publishing case studies on each of the 10 sites, and planned to share its findings with other hospitals through Children's Hospital Association meetings, academic conferences, and a webinar series. The awardee leaders expected the pace of dissemination to pick up as the original cooperative agreement period came to a close and results were again available to validate the positive impacts they thought they were seeing in Medicaid data.

D. Factors associated with progress toward implementing the SSR plan

The awardee and the individual sites said they expected two main factors (beyond those related directly to the payment model) to help support SSR of the CARE program. Leaders at the awardee and the sites told us that engaging medical and administrative leaders and practice networks had helped bring about a culture change in terms of valuing the importance of focusing on CMC and their needs—and respondents expected this culture change to endure. Site staff and leaders said that having a larger institution's resources, such as research and IT staff, would make it easier to continue many of the program's processes, such as conducting surveys with the patients' families.

However, the awardee also noted that a key barrier to scaling the program to serve more children was that some hospital executives lacked awareness of the program's broader applicability. Some sites found it difficult to convince leaders that the program's concepts and processes can be useful to all children, not just those with significant medical complexity.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in the National Association of Children's Hospitals' CARE program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicaid beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: National Health Care for the Homeless Council

April 18, 2018

Shivani Reddy (RTI International)
Noah Triplett (RTI International)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's HCIA R2 program

The National Health Care for the Homeless Council received HCIA R2 funding to support the program, Medical Respite Care for People Experiencing Homelessness. Medical respite care is defined as acute and post-acute medical care provided to homeless individuals who are not sick enough to be in a hospital but who are too sick to recover from a physical illness or injury

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

on the streets. The services provided by respite care centers for these individuals are growing rapidly across the nation, and they vary widely. In light of these conditions, National Health Care for the Homeless sought to develop and implement a standard set of respite care services to serve as a model for other programs. The awardee implemented its program in five sites in which there were existing respite care programs for homeless individuals. The components of the program were care management, patient engagement, and transitional care coordination (Table I.1). There were no significant design changes in Year 3.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Provide comprehensive respite care in a setting that is safe for patients by consistently supplying and tracking the delivery of three core services: (1) care management, (2) patient engagement, and (3) transitional care coordination.
Major innovation	Implemented a standardized model of respite care in five sites. The model was defined by the delivery of a core set of services (the program components described below) that represented recognized best practices in the field of respite care.
Program components	**Care management** included the following: (1) case management (that is, the provision of social services); (2) medication monitoring; (3) prevention efforts such as provision of tobacco cessation counseling and treatment and influenza vaccination; and (4) establishment of a medical clearance date, which was determined by a medical doctor and indicated when a patient's health stabilized. **Patient engagement** primarily involved motivational interviewing and education in order to give patients the resources and confidence they needed to set goals for managing their own health. **Transitional care** established a primary care provider who would address a patient's need for care. Staff gave the patient and the primary care provider updated health care information and arranged follow-up primary care appointments within 7 days and 30 days after the patient was discharged from a hospital.
Target population	Respite care patients age 18 or older who were homeless. By definition, admission to respite care indicated that the patient had an acute illness or injury. These patients were considered a high-risk, high-cost, underserved population. Most of the patients qualified for or were enrolled in Medicaid.
Theory of change/theory of action	Access to respite care after a hospitalization for an acute injury or illness would provide participants with personalized care management that promoted goal setting and self-management. In turn, participants could better manage their chronic conditions, resulting in the use of more preventive services and primary care, fewer ED visits and hospitalizations, and lower health care costs.
Payment model	Prospective payment system (PPS) payments, value-based payments, bundled or episode payments
Award amount	$2,673,476
Effective launch date	March 2, 2015
Program setting	Five medical respite care sites: (1) Edward Thomas House in Seattle, WA; (2) Hennepin County Health Care for the Homeless and Catholic Charities in Minneapolis, MN; (3) Central City Concern in Portland, OR; (4) Circle the City in Phoenix, AZ; and (5) Columbus House in New Haven, CT.
Market area	Urban
Market location	Cities

Table I.1 *(continued)*

Program characteristic	Description
Target outcomes	• Decrease in total expenditures • Decrease in hospital admissions • Decrease in ED visits • Decrease in hospital readmissions within 30 days • Increase in participant self-management of chronic conditions • Increase in participant understanding of care plan • Increase in participant adherence to prescribed medications • Increase in smoking cessation efforts • Increase in number of participants who receive vaccinations as recommended • Increase in linkages of participants with social services • Improvement in care coordination

ED = emergency department.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that National Health Care for the Homeless was partly successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on six factors. First, the awardee did not meet its enrollment goals due to delays in enrollment and staffing, as well as an unanticipated shortage of respite care beds on average across sites. Second, the awardee struggled to collect data and complete surveys as planned because non-HCIA R2–funded staff who delivered services were not trained to deliver all of the intervention components and found the survey tools confusing. Third, although new clinical staff helped expand service delivery and data managers were hired to collect and track data, the project would have benefitted from a project manager at each site and more clinical staff to deliver services. Fourth, despite these challenges, the awardee delivered some of the services as intended by the standardized respite care model for transitional care (primary care outpatient follow-up), case management (obtaining housing for participants), and preventive care (tobacco cessation). Fifth, the awardee was successful in hiring staff with a high level of buy-in to the project and had limited staff turnover. In addition, staff reported that project trainings enabled them to engage a patient population that generally mistrusts health care professionals. Finally, survey results indicated that staff felt the program had a positive effect on care delivery, their ability to respond in a timely way to participants' needs, and the efficiency of care provided to participants.

Impact evaluation. Due to the lack of a strong comparison group, we do not anticipate being able to conduct a rigorous impact analysis for the National Health Care for the Homeless program. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. Prior to and during the cooperative agreement, the implementing sites received reimbursement from current payers through a variety of payment models as well as additional funding from grants, foundations, or hospitals. As part of the cooperative agreement, all five sites also actively reached out to new payers so that the sites could continue operating after HCIA R2 funding ended. The awardee is collecting cost data and lessons learned from all sites' payment models to communicate the value of respite care to payers and create a shared understanding of the costs associated with providing standardized respite care services.

Sustainability plans. National Health Care for the Homeless continued its program at all five implementing sites after the cooperative agreement ended, through payments from Medicaid payers. The sites developed and implemented various payment approaches and secured additional funding through other sources. They planned to continue care management and transitional care, in some cases by maintaining the new staff or, more commonly, embedding the services into other staff workflows and electronic medical records (EMRs). However, some sites were considering reducing or modifying services as a way to reduce program costs. The awardee and participating sites demonstrated interest in scaling and replicating the program. One site added a new program in Year 3.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of non-clinicians on their perceptions of the effect of the program on care delivery. The non-clinician staff survey, which was fielded from July 2016 to October 2016, achieved a response rate of 50 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items with fewer than 11 respondents, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partly successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

National Health Care for the Homeless enrolled participants in its medical respite care program who (1) were eligible for and already admitted to respite care and (2) consented to join the HCIA R2 program, allowing staff to collect and track their health data.

Respite care enrollment. Common eligibility criteria across the five participating respite care programs included requirements that individuals be (1) age 18 or older; (2) in need of further clinical care but not hospital-level care for an acute illness; (3) able to carry out activities of daily living; and (4) at high risk of readmission to a hospital if they returned to their former living arrangement, which may be the street. Individual sites could have additional criteria, which most commonly involved substance abuse. Three sites required participants to abstain

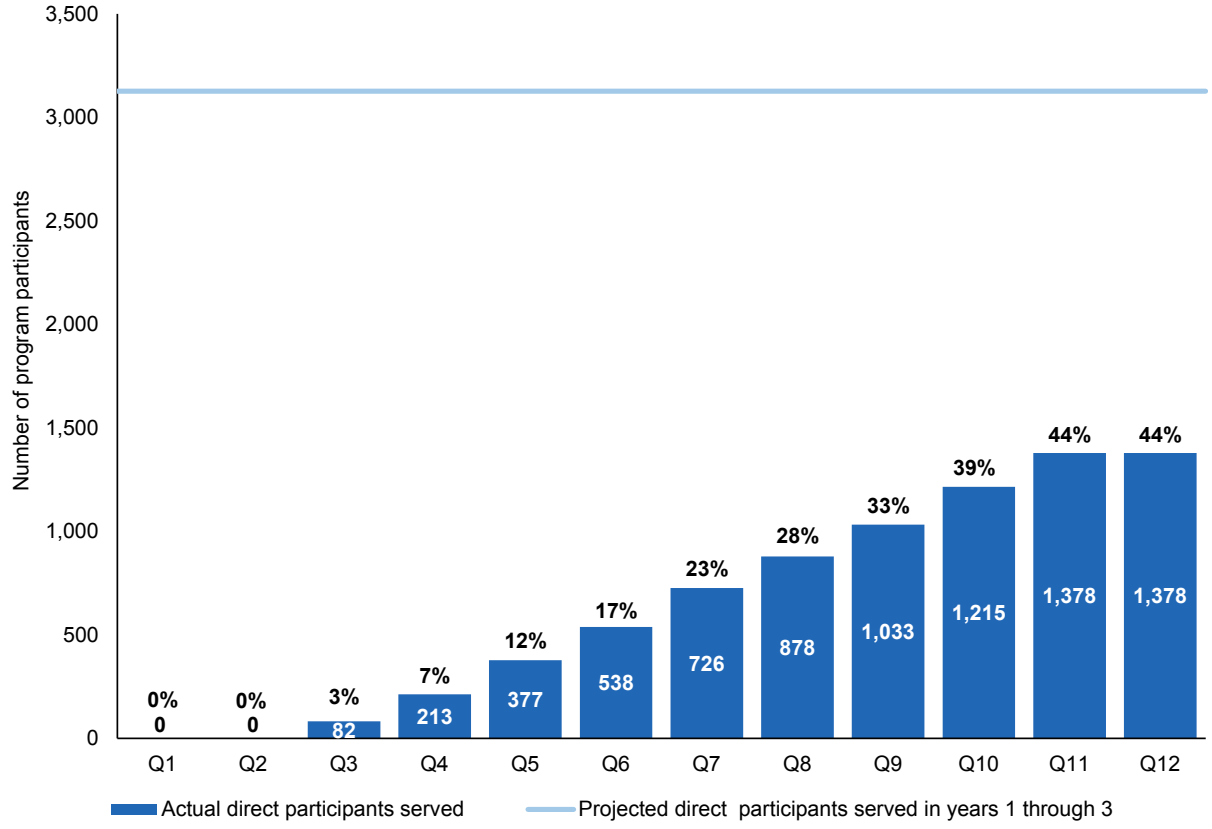
from alcohol or drug use, while two sites had harm reduction policies allowing active use of these substances. Although referral coordinators at sites could use a list of diagnoses to screen for eligibility, ultimately enrollment was determined on a case by case basis. Participants were referred to respite care centers primarily from hospitals, but also from other outpatient and community sites.

HCIA R2 program enrollment. Program staff approached individuals who were admitted to a respite care center about enrolling in the HCIA R2 program, explaining the program components and requesting consent to collect and track their health data. The data collected included information about hospitalizations and the respite care stay as well as administered surveys. In addition to the respite care program eligibility criteria, participants had to meet additional eligibility criteria to be eligible for the HCIA R2 program: individuals had to speak English and be cognitively able to consent, meaning that they were not impaired by severe mental illness or substance abuse. The program's target population and eligibility criteria did not change over time.

b. Evidence of enrollment effectiveness

Overall, National Health Care for the Homeless reported that it enrolled 1,378 participants from March 2, 2015, when it launched its program, through August 31, 2017—about 44 percent of its 3,127 three-year projected participants (Figure II.1). The awardee lowered its direct participant enrollment projection during the three-year cooperative agreement, from 4,039 at the beginning of the first program year to 3,127 in August 2015.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017)

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. National Health Care for the Homeless does not have indirect participants. Of note, the finder file received in September 2017 included a total of 1,441 participants due to a small number of enrollments after August 31, 2017. The awardee lowered its direct participant enrollment projection during the three-year cooperative agreement, from 4,039 at the beginning of the first program year to 3,127 in August 2015.

All five sites experienced lower-than-expected enrollment due to administrative delays, insufficient staffing, and longer-than-anticipated stays among the respite care participants. Several sites had institutional review boards and larger health system challenges that delayed enrollment by six months, which impacted enrollment numbers from the start. Project leaders described prolonged staff recruitment and training of the same duration. Furthermore, the pool of eligible program participants was drawn from patients already admitted to respite care centers. The average length of stay per participant exceeded 40 days, making it difficult to get new participants into respite care in a timely way and therefore reducing the number of eligible participants for the HCIA R2 program.

c. Barriers and facilitators associated with enrollment effectiveness

The awardee's inability to meet its three-year enrollment goal was influenced by two key barriers to enrollment into respite care. Two site-specific factors also facilitated or hindered enrollment at this stage. After admission to respite care, the awardee encountered one barrier to and one facilitator of enrollment in the HCIA R2 program.

Respite care enrollment. First, project leaders were challenged by administrative and staffing delays, which then delayed enrollment into respite care. Second, participants stayed in respite care much longer than expected, which led to a corresponding shortage of available beds and posed a significant barrier to enrollment. The awardee expected respite care stays to last two to four weeks. However, the average length of stay was as high as 49 days (seven weeks) during the cooperative agreement. Staff at respite care sites had several approaches to reducing the length of respite stay. One site expanded the number of respite beds by partnering with a local shelter; another site improved care transition processes to reduce the length of respite stay. A few sites attempted to reduce the number of inappropriate referrals of patients who were ineligible for respite care. Sites either performed quality improvement projects, conducted outreach, or gave presentations to hospitals and homeless shelters to educate referral sources on respite care, the referral process, and the eligibility criteria. However, only one site reported decreased length of stay as a result of these strategies.

In addition, two site-specific operations either facilitated or hindered enrollment: harm reduction and centralized respite care referrals. First, the two harm reduction sites admitted participants who were actively using alcohol or drugs. Some key informants felt that this allowed staff to meet participants where they were and continuously engage them until they decided that they were ready for services. However, other key informants noted that the volatile health status of these participants could be a barrier. Participants may be on the cusp of clinical improvement but quickly relapse into harmful health behaviors or clinically deteriorate, resulting in readmission to the hospital or self-discharge from respite care. Respite care sites addressed this barrier by waiting a few days to enroll a participant in order to select participants who were more likely to engage in respite care and the HCIA R2 program.

Second, a county policy change in Year 2 at one site required all referrals for persons experiencing homelessness to be triaged by a central, county-level staff member. Although more patients being discharged from hospitals were appropriately routed to respite care centers, community sites had trouble navigating the new system and some homeless patients may have slipped through the cracks.

HCIA R2 program enrollment. Staff reported concerns about participant confidentiality and wariness over joining a research study or government-funded program as key reasons that eligible participants declined to participate in the HCIA R2 program (though they were still enrolled in the respite care program). However, many key informants reported that overcoming participant mistrust greatly facilitated likelihood of consent. For example, multiple sites reported that using clinical or master's level–trained staff to help at the hospital referral stage facilitated the consent of participants into the program because a relationship with the participants had already begun. Some sites reported that having team meetings helped highlight successful enrollment strategies, such as identifying the best staff member to approach participants (that is,

a mental health specialist); adjusting workflows (for example, allowing staff to build relationships with participants for at least a few days before attempting to recruit them); and explicitly describing how the data would be used (that is, for the evaluation of the HCIA R2 respite program only).

2. Delivery of program services

a. Description of and changes to service delivery model

With HCIA R2 funding, National Health Care for the Homeless tried to establish a standardized set of services for respite care, including three core respite care services: (1) care management, (2) patient engagement, and (3) transitional care. The goal of delivering these three services was for participants to have greater confidence to manage their own health, engage in preventive care, and establish or strengthen the primary care they were receiving. Care management also included attention to other social determinants of health—particularly, securing housing for homeless patients—that can support patient engagement in health care. By shifting treatment of preventive, chronic, and ambulatory-sensitive acute care from inpatient settings (ED and hospital) to outpatient settings, National Health Care for the Homeless aimed to reduce health care costs and promote respite care as a defined service for future reimbursement.

Prior to the cooperative agreement, each site was providing inconsistent, varying levels of respite care services. Although the selected sites delivered many of these services prior to the cooperative agreement, implementing the standardized respite care model also provided a more formalized data collection process to help sites better track their work with participants and determine how to improve patient care, data collection, and reporting.

The implementation of this standardized set of respite care services was enabled by the hiring of additional staff, both clinical and non-clinical, who were supported by HCIA R2 funding. The respite sites had limited restrictions on the types of staff they could hire. Most sites opted for administrative or care coordination staff. In total, the HCIA R2 funding supported 11.75 full-time equivalent (FTE) staff, of which 5.01 FTE were dedicated to management or administrative staff and 2.97 FTE were for care coordinators, case managers, and patient navigators (Table II.2). The only role that the awardee required to be professionally licensed was the medical director, though sites did hire other licensed staff as well, such as social workers, mental health providers, and pharmacists. Data managers were a required position for each site. All sites hired a non-clinical data manager whose primary role was to collect and track the delivery of respite care services. Data managers used a standardized tool that aggregated data from multiple sources, including respite care records, surveys, and hospital data.

National Health Care for the Homeless provided training to staff to support implementation of the standardized respite care model. In Year 1 of the cooperative agreement, the awardee provided training webinars on patient-centered care, participant consent, and motivational interviewing to HCIA R2–funded staff. Staff who began after the first year of the cooperative agreement were given PowerPoint presentations and online resources to introduce them to the respite care model and to National Health Care for the Homeless.

Table II.2. Staff composition by FTE

Worker type	FTE new hires (Q1–Q12)
Care coordinator, case manager, or patient navigator	2.97
Management or administrative staff	5.01
Behavioral or mental health worker (non-physician)	1.00
Nurse, including registered nurses	1.33
Pharmacist	0.40
Physician assistant	0.50
Physician	0.00
Social worker	0.48

Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017)

FTE = full-time equivalent.

b. Evidence of service delivery effectiveness

National Health Care for the Homeless was partially effective in delivering program services. The awardee was most successful delivering respite care services related to transitional care coordination (primary care follow-up appointments), elements of care management (obtaining housing and supplemental income), and some preventive care (smoking cessation). The awardee was less successful in delivering other types of preventive care (influenza vaccinations). Despite initial delays, National Health Care for the Homeless met its staffing goals over the three-year period. In addition, staff at each of the five sites reported receiving effective training to work with the homeless population and implement intervention components. Staff were engaged in the program through a shared mission of caring for the homeless and a collaborative team approach. However, the awardee struggled to collect data and complete surveys as planned because non-HCIA R2–funded staff who delivered services did not receive training on the intervention components and found the survey tools confusing to use. Most new staff hires were administrative or non-licensed, which limited the delivery of services including vaccinations and prescription of smoking cessation treatments. Although data managers were hired to collect and track data, project leaders recognized that the project would have benefitted from a project manager (who also managed data) to direct implementation of the project at each site.

Delivery of intervention services. Within the framework of standardized respite care service delivery, sites had flexibility in how to achieve their goal. They did not have specific national policies from National Health Care for the Homeless to guide their program implementation. Sites varied in how they adjusted their workflow and how they used staff to engage participants, as well as how often staff interacted with participants. For example, a social worker at one site engaged participants in goal setting, while a nurse at another site fit goal setting into the nurse's existing workflow. One site used a part-time pharmacist to help participants with medication management, while another site used a visiting nurse service. Sites also used different strategies to track the provision of services by respite care staff. One site placed a checklist on the door of each participants' room to keep track of which services were pending delivery. In general, respite staff aimed for daily interactions with participants, but were

given flexibility to tailor interactions based on participant needs. A social worker may have met with a participant several times a day at the time of admission, but then tapered visits to two or three times a week by the end of the stay.

Table II.3 shows implementation measures for the three specific categories of respite care services that were captured by the data aggregation tool.

Table II.3. Measures of implementation

Core service	Measure
Care management	Tobacco cessation
	Influenza vaccination
	Benefits (Medicaid, Supplemental Security Income, Temporary Assistance for Needy Families, food stamps, and housing)
Patient engagement	Consumer Assessment of Healthcare Providers and Systems—Health Literacy Supplement
Transitional care	Care Transition Measure–3
	7-day and 30-day primary care follow-ups

Source: Data collected by National Health Care for the Homeless

Care management. The five respite care centers collected data on the following care management elements: tobacco cessation, influenza vaccination, and case management efforts to obtain social services for participants.

- **Tobacco cessation.** Among the 1,441 respite care participants, 984 participants (68 percent) were current smokers. Of current smokers, 964 participants were offered a smoking cessation intervention (98 percent), nearly meeting the awardee's goal of 100 percent. A third of smokers accepted the intervention (n = 326), which did not meet the awardee goal of 50 percent, but was comparable to national data of smoking cessation.[3] Within the subset of smokers who accepted an intervention, two-thirds received pharmacotherapy (nicotine replacement therapy or Wellbutrin) in line with smoking cessation guidelines—better than national data, which show that only one-third of smoking attempts use evidence-based methods to quit smoking.[4]
- **Influenza vaccination.** Influenza vaccination was assessed among the 685 respite care participants who presented during influenza season (October 1 to February 2). Nearly half of these participants had already been vaccinated prior to arriving at respite care—which was not surprising because the majority of participants were discharged to respite care from a

[3] Babb, Stephen, Ann Malarcher, Gillian Schauer, Kat Asman, and Ahmed Jamal. "Quitting Smoking Among Adults—United States, 2000–2015." *Morbidity and Mortality Weekly Report,* vol. 65, no. 52, January 6, 2017, pp. 1457–1464.

[4] Agency for Healthcare Research and Quality. "Clinical Guidelines for Prescribing Pharmacotherapy for Smoking Cessation." Content last reviewed December 2012. Available at http://www.ahrq.gov/professionals/clinicians-providers/guidelines-recommendations/tobacco/prescrib.html.
See also Babb and colleagues (2017).

health care setting. However, only 117 of 323 eligible unvaccinated participants received a flu shot while in respite care (36 percent).

- **Benefits.** Respite care center staff attempted to obtain benefits for participants such as Medicaid, Supplemental Security Income (SSI), Temporary Assistance for Needy Families (TANF), food stamps, and housing (Table II.4). For participants who did not already have such benefits, respite care staff would assist them in submitting applications based on their eligibility. In some cases, applications would be approved or denied while the participant was in respite care. For applications in process at the time of discharge, a continuity plan was put into place. Housing was the benefit most frequently approved while the participant was in respite care (32 percent). A major goal of the program was to improve health outcomes by securing housing for participants. Many participants already had Medicaid (93 percent) and food stamps (68 percent). The greatest number of applications had been submitted, were in process, or had received decisions for SSI benefits and housing benefits. Missing data on benefits ranged from 3 percent (Medicaid) to 20 percent (housing).

Table II.4. Benefit status of participants in respite care programs

	Medicaid		SSI		TANF		Food stamps		Housing	
Benefit status	N	%	N	%	N	%	N	%	N	%
Already has benefit	1,333	93	511	35	4	0	987	68	67	5
Application submitted	4	0	142	10	0	0	51	4	337	23
Approved while in medical respite	32	2	41	3	0	0	108	7	460	32
Denied while in medical respite	0	0	9	1	0	0	2	0	3	0
Application in process and continuity plan in place	5	0	197	14	0	0	35	2	174	12
Patient not eligible for benefit	26	2	403	28	1,399	97	121	8	35	2
Patient declined assistance	0	0	15	1	1	0	9	1	58	4
Missing	41	3	123	9	37	3	128	9	307	21

Source: National Health Care for the Homeless

SSI = Supplemental Security Income; TANF = Temporary Assistance for Needy Families.

Transitional care. Transitional care refers to establishing a linkage to a primary care provider to address the participant's needs. This was accomplished by providing the participant and primary care physician with updated health care information, arranging follow-up primary care appointments within 7 days and 30 days after hospital discharge, and surveying participants on their experiences with transitioning out of respite care prior to discharge from respite care. To evaluate participant experience with the transition from respite care to outpatient care, the awardee used a modified Care Transition Measure (the CTM-3) to survey participants prior to discharge from respite care, though sometimes staff completed these weeks prior to discharge to

ensure that the data were collected. The CTM-3 measures understanding of the role of self-care, medication management, and incorporation of participant preferences into the care plan.[5]

Results for the modified CTM-3 are reported in Table II.5. The mean CTM-3 score was 80, with a standard deviation (SD) of 21. Site-specific means ranged from 71 (SD 35) to 84 (SD 18), which corresponds to most participants reporting positive experiences with transitional care, though these are also sites that had a greater proportion of missing data. Studies conducted in similar participant populations—those with medical and mental health comorbidities or cognitive impairments—found similar mean CTM-3 scores of 79 to 81.[6]

Table II.5. Modified mean CTM-3 score for respite care participants through August 31, 2017

	CTM-3
	Mean (SD)
Total (n = 896 of 1,441)	80 (21)
Phoenix, AZ (n = 311 of 426)	83 (18)
New Haven, CT (n = 68 of 132)	75 (25)
Minneapolis, MN (n = 91 of 183)	71 (35)
Portland, OR (n = 333 of 445)	80 (18)
Seattle, WA (n = 93 of 255)	84 (16)

Source: National Health Care for the Homeless

Note: The CTM-3 questions were as follows: (1) The medical respite care staff took my preferences into account in deciding what my health care needs would be when I left the hospital; (2) When I left the medical respite program, I had a good understanding of the things I was responsible for in managing my health; and (3) When I left the medical respite program, I clearly understood the purpose for taking each of my medications. Survey respondent options for each question were strongly disagree, disagree, agree, and strongly agree. Each item was converted to a 0 to 100 score and the mean of the three items was computed (see National Quality Forum, "Specifications for the Three-Item Care Transition Measure—CTM-3," at https://mhdo.maine.gov/_pdf/NQF_CTM_3_%20Specs_FINAL.pdf). A lower score indicated poorer care transition.

CTM-3 = Care Transition Measure–3; SD = standard deviation.

National Health Care for the Homeless tracked primary care follow-up appointments after hospital discharge for the subset of patients who were discharged to respite care after a hospital

[5] Coleman, Eric A., Eldon Mahoney, and Carla Parry. "Assessing the Quality of Preparation for Posthospital Care from the Patient's Perspective: The Care Transitions Measure." *Medical Care,* vol. 43, no. 3, 2005, pp. 246–255.

Coleman, Eric A. "CTM Frequently Asked Questions." Denver, CO: University of Colorado Denver, 2016. Available at http://caretransitions.org/wp-content/uploads/2015/08/CTM_FAQs.pdf.

[6] Chan, Brian, L. Elizabeth Goldman, Urmimala Sarkar, Michelle Schneidermann, Eric Kessell, David Guzman, Jeff Critchfield, and Margot Kushel. "The Effect of a Care Transition Intervention on the Patient Experience of Older Multi-Lingual Adults in the Safety Net: Results of a Randomized Controlled Trial." *Journal of General Internal Medicine,* vol. 30, no. 12, 2015, pp. 1788–1794. doi:10.1007/s11606-015-3362-y.

Goldstein, Jennifer, LeRoi S. Hicks, Paul Kolm, William S. Weintraub, and Daniel J. Elliott. "Is the Care Transitions Measure Associated with Readmission Risk? Analysis from a Single Academic Center." *Journal of General Internal Medicine,* vol. 31, no. 7, 2016, pp. 732–738. doi:10.1007/s11606-016-3610-9.

admission (n = 899). There were no claims data for the majority of these patients, so respite admission dates were used as a proxy for hospital discharge dates. More than 65 percent of participants were seen in primary care within 7 days of admission to respite care and 96 percent of participants were seen within 30 days. Fourteen percent of hospitalized participants were missing data on primary care follow-up.

Staffing and training. The composition of teams varied by site as well as by responsibility; however, social workers and mental health specialists were typically responsible for all three components of the respite care model (care management, patient engagement, and care transitions). The role of nurses somewhat overlapped with that of social workers and mental health specialists, but often focused on care management responsibilities. Overall, key informants reported that hiring new staff enabled respite care sites to expand services, increase efficiency, and reach out to more participants. However, limited clinical staff may have hindered service delivery of some program components, particularly related to vaccination and smoking cessation.

Staff training. Although training webinars related to patient-centered care, participant consent, and motivational interviewing were not required, key informants reported that they were useful to incorporate new staff and help them develop confidence when interviewing participants. The awardee continued informal training throughout the cooperative agreement. All respondents to the survey of program staff reported participation in training via staff meetings, huddles, shadowing other staff, and self-study. A majority of respondents also reported mentoring and individual supervision. However, project leaders reported that expanding training to non-HCIA R2–funded staff and including a training on data collection could have reduced missing data and improved service delivery. Outside of the HCIA R2 project, the awardee provided ongoing technical assistance regarding working with homeless populations to help prevent provider burnout.

Recruitment and engagement of providers. Although there were delays in staffing, the awardee was careful to hire dedicated staff with the skill set to work with a homeless population and thus had limited turnover during the cooperative agreement. The awardee also undertook several strategies to engage and retain providers and provider organizations, including provision of technical assistance, performing site visits, and holding regular data manager calls, though these became less frequent over time. Within respite sites, staff prioritized communication and shared responsibility. In the survey of program staff, nearly all respondents agreed that members of the program team operated efficiently and collaborated effectively to meet participant needs. Likewise, most survey respondents reported that there was minimal resistance among staff to the intervention and that most staff had a shared understanding of program goals.

Some key informants reported that the medical director at their site played the role of program champion, which increased buy-in from other providers. However, project leaders said that placing a dedicated project director at each site could have facilitated implementation of service delivery and data collection. Data managers did not always have the authority or expertise to direct clinical and non-clinical staff to complete project components. In addition, clinical staff could lose sight of meeting program goals in busy clinical settings.

National Health Care for the Homeless developed and maintained relationships with partner organizations throughout the collaborative agreement. The awardee engaged provider organizations, most commonly the referring hospital, through presentations about the program and its objectives. In interviews, staff discussed the value of this approach in increasing buy-in and support from hospital staff, such as social workers or discharge planners. For some sites, ongoing communication between partnering hospitals included weekly or quarterly check-ins to help coordinate referrals and data sharing.

Engagement of program participants. The goal of patient engagement was to give participants the resources and confidence that they needed to set goals for managing their health. Key informants said that program staff had a strong understanding of the homeless population they served, which allowed them to better engage participants. Respite care staff were keenly aware of participants' fear of and skepticism toward medical professionals. Several staff who were interviewed said they had to figure out how to best gain participant trust. One strategy included using the "medical hospitality" approach—giving participants a participatory role in their health care and allowing development of relationships with the staff. Staff promoted engagement through frequent encounters with participants, sometimes aiming to increase the number of clinical staff so that participants could be seen at least every other day. The residential setting also enabled regular contact between staff members and participants, through one-on-one participant encounters, support group meetings, and educational sessions about tobacco cessation and access to benefits. Elements of the intervention, including motivational interviewing, goal setting, and participant education supported staff efforts to engage participants. Nearly all survey respondents reported successfully engaging participants in the program. Multiple interviewees noted the importance of having access to mental health services for participants. The presence of these services in respite care centers facilitated patient engagement and follow-up of mental health issues.

To measure the degree to which respite care participants had the capacity to process and understand health information, National Health Care for the Homeless surveyed participants prior to discharge from respite care by using five modified questions from the Health Literacy supplement of the Consumer Assessment of Healthcare Providers and Systems (CAHPS-HL) survey. This subset of questions is recommended by the Agency for Healthcare Research and Quality to capture a global perspective on health literacy.[7] The mean composite CAHPS-HL score for all participants and by state are shown in Table II.6. Missing data for the sites ranged from 24 percent (Oregon) to 63 percent (Washington). Overall, 37 percent of participants were missing CAHPS data.

[7] Agency for Healthcare Research and Quality. "About the CAHPS Item Set for Addressing Health Literacy." 2014. Available at https://www.ahrq.gov/cahps/surveys-guidance/item-sets/literacy/index.html.

Weidmer, Beverly A., Cindy Brach, and Ron D. Hays. "Development and Evaluation of CAHPS® Survey Items Assessing How Well Healthcare Providers Address Health Literacy." *Medical Care,* vol. 50, September 2012, pp. S3–S11. doi:10.1097/MLR.0b013e3182652482.

Table II.6. CAHPS-HL composite scores, overall and by state

	CAHPS composite
	Mean (SD)
Total (n = 423 of 1,441)	82 (21)
Phoenix, AZ (n = 312 of 426)	82 (21)
New Haven, CT (n = 69 of 132)	77 (24)
Minneapolis, MN (n = 92 of 185)	89 (16)
Portland, OR (n = 338 of 445)	80 (22)
Seattle, WA (n = 95 of 255)	87 (17)

Source: National Health Care for the Homeless

Note: The modified CAHPS-HL questions are as follows: (1) During your stay, how often did this provider give you all the information you wanted about your health? (2) During your stay, how often did this provider encourage you to talk about all your health questions or concerns? (3) During your stay, how often did this provider ask you to describe how you were going to follow self-care instructions? (4) If you received test results during your stay, how often were the results of your blood test, x-ray, or other test easy to understand? (5) During your stay, how often were instructions about how to take your medicines easy to understand? Survey respondent options for each question were always, usually, sometimes, and never. As recommended by Weidmer and colleagues (2012), the mean of each question was transformed to a 0 to 100 score and the composite CAHPS-HL score was computed from the mean of the five questions.

CAHPS = Consumer Assessment of Healthcare Providers and Systems–Health Literacy; SD = standard deviation.

The mean CAHPS-HL score for the total sample of participants was 82 (SD 21), which corresponds to the majority of participants indicating that they usually or always have an understanding about health questions and is comparable with prior published data for this measure.[8] This score met the awardee's goal. However, given the significant level of missing data (37 percent)—particularly in the sites that launched later than the others in Year 1 (63 percent in Seattle and 50 percent in Minneapolis)—higher scores may reflect nonresponse bias and must be interpreted with caution.

c. Barriers and facilitators associated with service delivery effectiveness

The awardee's ability to effectively deliver services was influenced by several factors. While reviewing the full three-year cooperative agreement, key informants identified three key facilitators of service delivery: (1) committed staff working in a collaborative environment, (2) trainings on patient engagement, and (3) structural features that facilitated transitional care.

First, National Health Care for the Homeless hired dedicated staff with the skill set to work with a homeless population, which fostered collaboration. Staff reported shared workloads and productive communication between staff members. Key informants described how huddles (regular check-ins with the entire team) kept staff members engaged in the intervention, allowed them to receive feedback, and provided updates on different components of the intervention. Staff said that limited turnover was a result of feeling like they contributed to improved participant outcomes. This was supported by the program's monitoring data, which showed that there were few staff separations during the cooperative agreement. Although the data manager

[8] See Weidmer and colleagues (2012).

was a new role, multiple key informants reported that the data managers could identify gaps in care, which facilitated the consistent delivery of a wider range of services within respite care. For example, staff members reported more consistently querying participants about influenza vaccination and smoking cessation. Other examples of changes included the introduction of regular interdisciplinary team meetings and a standardized discharge process. New staff allowed sites to expand their services, reach more homeless individuals, and focus on the provision of care. Some staff expressed a desire for more communication and collaboration between respite care sites; they suggested a platform for sharing lessons learned and best practices across respite care sites.

> "I think [the] philosophy of all of these projects is to make sure that they hire people who basically have a nonjudgmental attitude, who understand some of the root causes of homelessness, understand how all of the different parts of the system connect with each other."
>
> —Program leader

Second, trainings were viewed favorably by staff, facilitating staff integration into the project, helping them better understand the goals of the project and teaching them new skills. Respondents described how learning motivational interviewing techniques helped mitigate institutional distrust and foster patient engagement. Others reflected on how the goal-setting exercises gave participants an opportunity to express their own needs and explore barriers to meeting their goals. In the survey of program staff, all respondents felt the trainings provided helpful skills and improved job performance. Fewer than half of the staff survey respondents desired additional training to help meet job responsibilities.

> "Motivation[al] interviewing … opened doors to be able to talk to clients.… They can share on their own terms what it is that they need help with.… To say, this is what I need, this is what's going on. So, the motivation[al] interview really helps in terms of communication with clients and help with that."
>
> —Case manager

Third, structural features, such as physical location of or type of EMR at the respite care sites facilitated outpatient primary care follow-up within 30 days of admission to respite care. Transitional care at different sites was facilitated by proximity to or colocation of primary care offices, long-standing partnerships with the primary care community, or a shared EMR system. Care managers in some sites physically escorted participants to their follow-up appointments to ensure that this standard of respite care was met.

Program leaders and staff also described three significant barriers to the delivery of program services: (1) inadequate staffing, (2) need for broader training, and (3) a challenging participant population.

First, although new staff members were dedicated to the awardee's mission of caring for the homeless, the specific types of staff hired may not have been adequate to meet service delivery goals. Project leaders reported that the addition of a program manager with broader authority over implementation, rather than the narrowly defined data manager role, may have improved implementation of service delivery and, accordingly, data collection. The limited availability of licensed independent medical providers constrained the ability of the awardee to offer first-line pharmacotherapy for smoking cessation and may have also impacted delivery of influenza vaccinations. A greater number of participants were evaluated for various social service benefits, though not all sites had a dedicated case manager. For sites that did have a case manager, that

staff member may have had multiple duties beyond traditional case work (for example, substance abuse counseling).

Second, project leaders noted the challenges with data collection and suggested that they should have offered training to meet this goal. Significant amounts of missing data for implementation measures, particularly for the CAHPS-HL and CTM-3 scores, underscored the substantial barriers to accurately tracking and collecting data. Sites had different constraints on collecting and reporting data, such as who was eligible to collect data, how often those staff could be on-site, and confusion over how to use the data aggregation tool. Key informants expressed the desire for a webinar on this topic. Some data managers reached out to one another across sites to try to share lessons learned. However, missing data continued to be a challenge throughout the cooperative agreement. Furthermore, all clinical and non-clinical staff were involved with service delivery, but only HCIA R2–funded staff received training modules. Interviewees reflected that expanding trainings to all staff involved with service delivery, regardless of funding, would have improved training efforts.

Third, the complexity of the participants' circumstances prevented some participants from becoming fully engaged in the program, while also creating some participant mistrust. Participants with a volatile health status may relapse and have to return to the hospital or leave against advice before engaging in respite care services. Homeless individuals' mistrust of health care professionals and institutions was also a significant barrier to patient engagement that was repeatedly reported by key informants over the course of the cooperative agreement. This also may explain the limited response rate to the survey tools. As previously discussed, staff tried to mitigate participant mistrust by developing a relationship with the participant prior to approaching him or her about enrolling in the program.

C. Assessment of perceived program effects on the delivery of care and outcomes

1. Data and methods for assessing perceived program effects on care

We used survey data to describe the experiences of non-clinician staff involved in the program as well as their perceptions of the program's effects on care delivery and health outcomes. The non-clinician staff survey, which was fielded in fall 2016, achieved a response rate of 50 percent. All survey items had fewer than 12 respondents, so we describe the findings in qualitative terms. We supplemented the survey findings with data collected through the site visit interviews.

2. Overall assessment of perceived program effects on care

Overall, staff perceived the effects of the intervention in a positive way. In the staff survey, most respondents felt the program was making a difference in meeting critical needs in the community and positively impacting participants' health goals. All staff respondents felt the program had been effective in achieving its goals.

3. Review of factors associated with perceived program effects on care

Program leaders and staff cited several factors as contributing to their perceptions of the positive effects of the program on care delivery and health outcomes. Internal analyses

conducted by the awardee suggested a reduction in the total cost of care, hospital admissions, and ED visits in two of the three sites with available Medicaid claims data.[9] In the survey of program staff, all respondents agreed that the program had a positive impact on the quality of care and services they provided to participants. All respondents also indicated that they felt the program positively impacted participant satisfaction and quality of life. Smoking cessation efforts were considered an example of the program's positive impact on the delivery of care. For example, querying smoking status led to increased communication among the medical team. This finding was complemented by staff survey data, in which all respondents indicated that the program positively impacted care coordination. All staff respondents indicated that the program positively impacted (1) their ability to respond in a timely way to participant needs and (2) the efficiency of care or services they provided to participants.

Interviewees reported other positive effects of the program, including much-needed attention to missing data and inconsistent care practices. This information helped clinical staff take specific actions to improve patient care processes and access. In the survey of program staff, nearly all respondents indicated that the program positively impacted access to care or services for all participants.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

The difficulties reported with enrollment and patient engagement underscored the challenges of working with this vulnerable population. These challenges also limited sample size and our ability to detect program impacts. The main effects of the program were expected to be a reduction in ED visits, hospitalizations, and readmissions, leading to a decrease in total cost of care. Based on the awardee's theory of action, the program attempted to increase outpatient utilization. If participants effectively transitioned to outpatient care, then it would be reasonable to observe a change in core outcomes within a year. The program also attempted to improve health care outcomes by obtaining housing benefits, which may further decrease costs. The program may yield varying outcomes in sites with EMRs, which may facilitate transition to outpatient care, as well as sites with harm reduction policies. Sites that use harm reduction accept substance abusing participants who may benefit more greatly from the program due to their vulnerable health status, though the chronic relapsing nature of addiction may preclude positive findings within a year of follow-up.

[9] The awardee presented a pre-post analysis of combined Medicaid and Medicare claims data in its 12th quarter narrative report. It was unclear whether or not regression analysis was performed.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of the National Health Care for the Homeless's program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on the three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

The National Health Care for the Homeless program ended August 31, 2017. The projected sample sizes in Table III.1 are based upon actual enrollment as of February 28, 2017, and they are the maximum number of beneficiaries that could be included in our evaluation to allow for all participants to receive six months of program exposure, a requirement for inclusion in our evaluation to minimize dilution of program effects due to limited exposure to the program components. Due to processing lags in Medicaid data, we have not confirmed that all 656 Medicaid beneficiaries meet program eligibility for inclusion in our impact evaluation.

Table III.1. Assessment of HCIA R2 awardee evaluability as of August 31, 2017: National Health Care for the Homeless

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure	107[a]
Projected Medicaid population with 6 months of program exposure	656[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,003
Likelihood of all-cause hospitalizations	161
MDE sample size requirement to detect 20% effect	
Total expenditures	251
Likelihood of all-cause hospitalizations	40
Participation/Selection bias of concern	Yes, participant self-selection high/high refusal rate
Intervention implemented that differs from baseline period	Questionable, patients may have been receiving intervention prior to HCIA R2 cooperative agreement
Claims sufficient to identify treatment and comparable comparison group?	Questionable, no testing yet to determine strength of intent-to-treat framework
Likelihood of solid comparison group	Serious concern. We may be not able to identify a strong comparison group

Table III.1 *(continued)*

Evaluability domain	Response
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Lack of strong comparison group
Survey data for treatment group that will be analyzed	Staff survey
Implementation data that will be analyzed	Implementation data provided by awardee for treatment group that will allow us to examine implementation of the core elements of respite care—care management, patient engagement, and transitional care

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

At this point, we do not expect to conduct a rigorous impact evaluation of the awardee's program. First, the awardee targeted a vulnerable population that may not have had insurance until individuals were exposed to the intervention, so we are lacking baseline data. Second, there are significant challenges involved in using claims data to identify a solid comparison group composed of homeless individuals. Third, our preliminary approach to identifying a comparison group is based on a hospital discharge, but nearly 40 percent of respite care patients have been admitted from a community setting and there is no triggering event to use to identify a comparison group of beneficiaries.

We reviewed Medicaid data for program participants in Arizona and Oregon, which included records from their program start date through September 2016 and included one year of baseline data. We examined the proportion of participants with baseline data and explored the feasibility of using ICD supplemental diagnosis codes reflecting homelessness.[10]

Among the 345 participants in Arizona, 236 of them (68 percent) had baseline data, but only 61 of the 236 had an ICD code indicating homelessness. In Oregon, a greater proportion of Medicaid beneficiaries had baseline data (210 of 236 participants, or 89 percent), although only 125 participants were discharged from a hospital (59 percent) and had a claim with an ICD code indicating homelessness. Furthermore, comparing the top 10 principal diagnoses in participants with and without homelessness codes yielded only three overlapping diagnoses in each state, indicating that these groups are not comparable. Results from a subset of participants with homelessness codes would not necessarily be generalizable to the whole sample of participants receiving the intervention.

We are currently exploring Medicaid data received from the Connecticut site for its 132 participants. However, previous internal analyses by the awardee indicated that homelessness

[10] A diagnosis of homelessness is designated by the following ICD-9 codes: lack of housing (V60.0), inadequate housing (V60.1), other housing or economic circumstances (V60.89), or unspecified housing or economic circumstances (V60.9). V codes are used to describe beneficiaries' encounters with circumstances other than disease or injury, including circumstances that can influence a person's health status, such as disease exposures or, in our case, homelessness. The corresponding ICD-10 codes are as follows: homelessness (Z59.0); inadequate housing (V59.1); other housing or economic circumstances (V59.8); or problems related to housing and economic circumstances, unspecified (V59.9).

ICD supplemental diagnosis codes were used for only 25 percent of respite care patients discharged from the hospital. Thus, we do not expect more than about 30 potential participants for analysis. Likewise, restricting the smaller subset of Medicare participants with an inconsistently used ICD diagnosis code will likely yield a sample too small to conduct an impact analysis.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents our summary of the baseline characteristics of the treatment group, which we measured during the 12 months before each beneficiary's enrollment date and the initiation of respite care services. The treatment group being assessed for this evaluation consisted of Medicaid and Medicare fee-for-service (FFS) beneficiaries who were enrolled in the program. The majority of participants were enrolled in Medicaid. We have received state Medicaid data from the awardee for the treatment group participants in a limited number of states as described in Section III.A.

National Health Care for the Homeless enrolled participants in the five respite care program locations as follows: Phoenix, Arizona, and New Haven, Connecticut, starting on March 2, 2015; Portland, Oregon, starting on June 8, 2015; Seattle, Washington, starting on August 17, 2015; and Minneapolis, Minnesota, starting on September 21, 2015. As of May 2016, the awardee had enrolled 748 participants in the program, including 603 participants with Medicaid-only benefits.[11] The remaining participants had Medicare, were dual eligibles, or were uninsured.

In presenting the baseline characteristics in this report, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who had met all criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they had to be enrolled in the awardee's program on or before May 31, 2016, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. The calendar period covered by the baseline quarters was based on the enrollment date for each enrollee and therefore varied by enrollee. After we excluded beneficiaries who did not meet the above criteria, a total of 61 participants (8.2 percent) were included in the analysis of baseline characteristics for this report.

The Medicare FFS beneficiaries participating in respite care were characterized by a high level of disability and medical comorbidity (Table III.2). The majority of participants in the treatment group were younger than 65 (77 percent), male (90 percent), and white (75 percent).

[11] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

The original reason for Medicare eligibility was primarily disability (79 percent), compared with 24 percent of Medicare beneficiaries nationwide. Only 16 percent of participants were originally eligible for Medicare because of age or survivor's insurance, while 5 percent of participants were eligible because of end-stage renal disease (ESRD). Seventy-nine percent of participants were dually eligible for Medicare and Medicaid, although in some cases Medicaid benefits may have been restricted to the payment of co-insurance and deductibles. The average hierarchical condition categories (HCC) risk score of participants was 2.7, which meant that patients recruited for the respite care programs were predicted to be 170 percent more costly than the general Medicare FFS population in the first year of the program.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in National Health Care for the Homeless's program through May 31, 2016

	All participants (N = 61)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	47	77
65 to 74	14	23
75 to 84	0	0
85 and older	0	0
Gender		
Female	6	10
Male	55	90
Race		
White	46	75
Black	11	18
American Indian, Alaska Native, Asian/Pacific Island American, or other	2	3
Hispanic	1	2
Original reason for Medicare eligibility		
Old age and survivor's insurance	10	16
Disability insurance benefits	48	79
ESRD[a]	3	5
Hospice[b]		
Medicare/Medicaid dual status, percent dual[b]	48	79
HCC score[c]		Statistic
Mean		2.7
25th percentile		1.48
Median		2.33
75th percentile		3.07

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Table III.2 *(continued)*

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the awardee indicated the beneficiary was enrolled in the respite care program. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition categories.

In Table III.3, we report baseline utilization and expenditure data for the treatment group on a common set of measures, including the four core measures from CMMI. The awardee attempted to lower the total cost of care by reducing ED visits, hospital readmissions, and hospital length of stay. We examined baseline cost of care by calculating average per beneficiary per month (PBPM)[12] Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $4,357 compared to the national average Medicare spending of $864 PBPM.[13] The quarterly PBPM payments ranged from $2,840 to $7,969. The average PBPM Medicare payment for inpatient services ($2,516) was the largest driver of the total cost of care and was comparable to or less than inpatient costs observed in studies of other urban homeless populations, such as in Harris County, Texas ($4,023 PBPM)[14] and Chicago, Illinois ($2,085 PBPM).[15] Quarterly expenditures for inpatient services in the final baseline quarter were two to four times those of previous quarters, which we would anticipate in an intervention that admits many participants following a hospital discharge.

[12] The months referred to in our calculations are 30-day periods rather than calendar months.

[13] Kaiser Family Foundation, "Medicare Spending Per Enrollee, by State." Available at http://kff.org/medicare/state-indicator/per-enrollee-spending-by-residence/. Accessed October 25, 2016.

[14] Buck, David S., Carlie A. Brown, Karoline Mortensen, John W. Riggs, and Luisa Franzini. "Comparing Homeless and Domiciled Patients' Utilization of the Harris County, Texas Public Hospital System." *Journal of Health Care for the Poor and Underserved,* vol. 23, no. 4, 2012, pp. 1660–1670. doi:10.1353/hpu.2012.0171.

[15] Basu, Anirban, Romina Kee, David Buchanan, and Laura S. Sadowski. "Comparative Cost Analysis of Housing and Case Management Program for Chronically Ill Homeless Adults Compared to Usual Care." *Health Services Research,* vol. 47, no. 1, pt. 2, February 2012, pp. 523–543. doi:10.1111/j.1475-6773.2011.01350.x.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in National Health Care for the Homeless's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	61	54	57	60	61
Average Medicare expenditures PBPM[a]					
Total	4,357 (655)	2,887 (782)	2,840 (883)	3,648 (856)	7,696 (850)
Acute inpatient	2,516 (434)	1,423 (454)	1,536 (708)	2,075 (707)	4,783 (617)
Inpatient other[b]	175 (73)	255 (213)	79 (76)	230 (194)	137 (93)
Outpatient[c]	644 (124)	491 (115)	515 (121)	533 (129)	1,001 (205)
Physician services	630 (100)	412 (89)	384 (93)	553 (119)	1,115 (195)
Home health	48 (15)	11 (11)	25 (17)	47 (24)	104 (40)
Skilled nursing facility	318 (104)	272 (160)	285 (181)	192 (114)	508 (182)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	27 (8)	23 (9)	18 (8)	18 (11)	47 (26)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	2,296 (386)	1,502 (534)	1,172 (390)	1,997 (612)	4,292 (478)
Outpatient ED visits	7,117 (1,496)	4,430 (1,053)	4,248 (815)	5,923 (1,579)	13,208 (3,638)
Observation stays	795 (320)	225 (126)	732 (289)	895 (395)	1,255 (692)
Primary care visits in any setting	20,539 (3,102)	16,294 (3,882)	15,161 (3,413)	17,770 (3,765)	31,764 (4,362)
Primary care visits in ambulatory settings	8,283 (1,311)	6,683 (1,198)	8,130 (2,101)	6,543 (1,332)	11,491 (1,766)
Specialist visits in any setting	25,660 (3,763)	15,693 (3,301)	15,234 (3,671)	26,585 (5,804)	42,925 (6,941)
Specialist visits in ambulatory settings	7,576 (1,151)	6,758 (1,565)	4,101 (873)	6,887 (1,600)	12,085 (2,008)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	82 (5)	19 (5)	16 (5)	26 (6)	72 (6)
Percentage with an outpatient ED visit[e]	87 (4)	45 (7)	52 (7)	52 (7)	73 (6)
Percentage with an observation stay[f]	34 (6)	6 (3)	13 (4)	14 (4)	13 (4)
Percentage with a 30-day readmission among all discharges	38 (5)	48 (10)	22 (10)	37 (13)	37 (7)
Percentage of participants with a readmission among all participants	24 (6)	11 (4)	7 (3)	5 (3)	17 (5)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

The Medicare FFS beneficiaries enrolled in respite care were high utilizers of acute care services. The rate of acute care hospitalizations was 2,296 per 1,000 Medicare FFS participants per year during the baseline year, compared with the national rate of 274 acute care hospitalizations per 1,000 Medicare FFS beneficiaries.[16] Other studies have reported similar

[16] See National Health Care for the Homeless Council. "What is Medical Respite Care?" Available at https://www.nhchc.org/resources/clinical/medical-respite/.

rates of hospitalization in homeless populations (3,000 to 3,600 hospitalizations per 1,000 beneficiaries).[17,18] Eighty-two percent of participants had at least one hospitalization during the 365 days before enrollment. The rate of ED visits was 7,117 per 1,000 participants per year in the baseline year, compared with a national rate of 652 per 1,000 Medicare beneficiaries annually.[19] The rate of ambulatory observation stays was 795 per 1,000 beneficiaries per year in the baseline year. Overall, observation stays at baseline far exceeded the national average of 58 per 1,000 beneficiaries in 2014.[20] For all acute care services, fourth quarter utilization rates were nearly double those of previous quarters, which may represent the transfer of patients from an ED or hospital observation unit to respite care rather than inpatient admission.

In the baseline year, the rate of primary care visits (8,283 per 1,000 Medicare FFS participants per year) was similar to the rate of specialty services in ambulatory settings (7,576 per 1,000 Medicare FFS participants per year). Primary care and specialist visits in ambulatory settings were strikingly high in the fourth quarter. This may be a result of some participants being referred to respite care from clinical outpatient sites, given that approximately 40 percent of all participants (not just Medicare FFS beneficiaries) were recruited from non-acute care settings. However, we may be observing a selection bias of high utilizers in an insured homeless sample. The odds of an insured homeless patient using ambulatory care services are two and a half times higher than those of a homeless patient who lacks insurance.[21] In addition, a Canadian study suggested that health care utilization outliers were more prevalent among homeless patients compared with controls who were not homeless.[22]

Thirty-eight percent of all hospital discharges were followed by a readmission in the 30-day post-discharge window in the baseline year. However, we could be observing an artificially low readmission rate in the fourth quarter if patients who enrolled after an admission at the end of the baseline period were potentially protected from readmission by the respite care intervention. The percentage of hospital discharges with a 30-day readmission (38 percent) was approximately twice the national average for Medicare beneficiaries (18 percent).

[17] See Buck and colleagues (2012).

[18] Sadowski, Laura S., Romina A. Kee, Tyler J. VanderWeele, and David Buchanan. "Effect of a Housing and Case Management Program on Emergency Department Visits and Hospitalizations Among Chronically Ill Homeless Adults: A Randomized Trial." *JAMA,* vol. 301, no. 17, 2009, pp. 1771–1778.

[19] See the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed October 2016.

[20] See the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2016. Available at http://medpac.gov/-documents-/data-book. Accessed October 2016.

[21] Kushel, M. B., E. Vittinghoff, and J. S. Haas. "Factors Associated with the Health Care Utilization of Homeless Persons." *JAMA,* vol. 285, no. 2, January 10, 2001, pp. 200–206.

[22] Hwang, S. W., and M. J. Henderson. "Health Care Utilization in Homeless People: Translating Research into Policy and Practice." Rockville, MD: Agency for Healthcare Research and Quality, October 2010. Available at http://gold.ahrq.gov.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

The five implementing sites of National Health Care for the Homeless received reimbursement from existing payers and additional funding from grants, foundations, or hospitals prior to and during the cooperative agreement. They were also actively reaching out to new payers to continue the respite care program, with some modifications, after HCIA R2 funding ends. The awardee is collecting cost data and lessons learned from all sites' payment models to communicate the value of respite care to payers and create a shared understanding of the costs associated with providing standardized respite care services.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

National Health Care for the Homeless provided cost analysis and leadership support to its five sites as they developed and implemented a multiple-payer payment model that was built on existing payment models. The five sites' payment models varied by payer, payment approach, and additional funding streams, primarily because each site offered different clinical services and was subject to state Medicaid policies that varied by states. Because the sites' clinical capacity to provide respite services varied, sites' payment models were categorized in the following two ways: (1) as a site affiliated with a federally qualified health center (FQHC) or (2) as a site affiliated with health care providers who were not part of an FQHC. Within each payment model category, different Medicaid-contracted and Medicare-contracted payers exist that utilize different payment approaches.

- **Three sites affiliated with an FQHC: Oregon, Arizona, and Minnesota.** Sites in Oregon and Arizona have been receiving payments for primary care encounters and respite care services for Medicaid FFS enrollees at their affiliated FQHCs through an enhanced prospective payment system (PPS) rate from the state Medicaid agency. They also receive additional payments for respite care services through Medicaid-contracted managed care organization (MCO) agreements for Medicaid managed care enrollees. The Arizona site contracts with six Medicaid MCOs and receives payments for respite care services through a state-established PPS rate, while the Oregon site receives payments from two MCOs through bundled payment mechanisms with a payment amount specific to MCOs. The Minnesota site's FQHC receives a flat monthly rate payment per beneficiary for respite care services provided via a Medicaid- and Medicare-contracted accountable care organization (ACO).
- **Two sites not affiliated with an FQHC: Connecticut and Washington.** The Washington site, which has had four Medicaid-contracted MCO agreements, receives payments for respite care services through a daily rate arrangement. The Connecticut site is not a medical facility and thus contracts with a visiting nurse program, which receives FFS payments for respite care services from the patient's medical provider, mainly the Medicaid state agency.

Table IV.1. Payment models by program sites

Site	FQHC affiliation	Payer	Payment
Central City Concern (OR)	Yes	MCO Medicaid	FFS
		State Medicaid	FQHC PPS
		Hospital systems, local government	FFS
Circle the City (AZ)	Yes	MCO Medicaid	FQHC PPS
		State Medicaid	FQHC PPS
		Hospital systems, hospice	FFS
Catholic Charities/Hennepin County Health Care for the Homeless (MN)	Yes	ACO (Medicaid/Medicare)	FFS
Edward Thomas House (WA)	No	MCO Medicaid	FFS
Columbus House (CT)—external visiting nurse program	No	Patient's medical provider	FFS

Source: Program sites.

ACO = accountable care organization; FFS = fee-for-service; FQHC = federally qualified health center; MCO = managed care organization; PPS = prospective payment system.

Some respite care services and operational expenses, such as staff time used for patient engagement, were not covered by the payers. Therefore, all sites also received additional funding from grants, foundations, and partnering hospitals. For two sites, partnering hospitals pay for a block of beds based on costs associated with those beds per night. Typically, hospitals are the first to engage with respite care centers in a payment arrangement because they can gain the most savings by shortening the index length of stay and avoiding future readmissions and ED visits. However, as sites experience an increase in MCO agreements, key informants expressed concerns that their hospital partners may become reluctant to pay or would initiate payment cutbacks because more services are now covered by Medicaid. National Health Care for the Homeless is working with these sites to communicate the value of respite care to their hospital partners in hopes of sustaining their funding. At the end of the cooperative agreement, hospital payments were still intact at pre-agreement levels. Sites are exploring a few potential solutions such as applying for hospital community benefit grants to support several non-billable services or setting up a membership fee in order for hospitals to refer patients.

> "This program more than anything helped create a framework to enroll people longitudinally in the study and have academically validated results. It has been tremendously valuable and has catalyzed our operating position and our dialogue with our local Medicaid agency and the MCOs that administer the plan."
>
> —Project lead at one of the five sites

C. Status of the payment model

The five program sites have all been receiving reimbursement from current payers and additional funding from grants, foundations, or hospitals. They were also actively reaching out to new payers. This will allow each of the sites to continue operating after HCIA R2 funding ends. Although many of these payment approaches existed prior to the cooperative agreement, they did not fully cover respite care costs. National Health Care for the Homeless has been providing technical assistance to help sites estimate the cost of the programs and facilitate negotiations with payers. Given differences in site structure and state policies, it is challenging to find

commonalities across sites and determine a standardized cost per service. To address this challenge, National Health Care for the Homeless is collecting data on the cost of providing services and lessons learned from all sites' payment models to communicate the value of respite care to payers and create a shared understanding of the costs associated with providing standardized respite care services.

D. Factors associated with the development of the payment model

Several sites that partner with MCOs have had challenges with establishing payments for respite care services. Typically, MCOs contract for a specific number of beds and that contract is based on a cost per month for each bed. National Health Care for the Homeless has not yet determined a per night cost for respite care services across the five sites due to the varying levels of clinical and non-clinical care capacity as well as the lack of a counterfactual for comparison. Without a standardized cost that respite care programs can present to MCOs for the value of their services, it is difficult for sites to know the appropriate level of reimbursement when trying to partner with MCOs. The lack of a standardized cost also means that one site can have a variety of different MCO contracts that cover different services at varying payment levels. Furthermore, MCOs' nondisclosure agreements do not allow National Health Care for the Homeless to know which services are included in the value-based or bundled payment codes. Not knowing which services constitute respite care or the appropriate reimbursement for those services, the awardee cannot easily gather best practices or build on existing MCO agreement successes.

To address these challenges, the awardee is working with sites to standardize respite care services as much as possible and to build a value proposition for MCOs to reimburse clinical providers for respite care services. National Health Care for the Homeless recently obtained Medicaid data from two state Medicaid agencies, which will facilitate internal return-on-investment analyses and can be used by sites in negotiations with payers.

One important lesson shared by a project lead at one of the program sites is to identify key stakeholders in providing care for the homeless and get buy-in from the community. Besides contracts with MCOs, some sites are actively developing value-based reimbursement strategies and having discussions with existing payers regarding shared savings. They also receive support from champions in the community and have been covering non-billable components, such as residential services, through philanthropy. In the absence of respite care, the local community and organizations will be responsible for the cost of post-acute care for homeless individuals. The awardee wants to convey the value of respite care to promote investment in this innovative outcome-driven program, rather than paying for the poor outcome in the absence of respite care.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

National Health Care for the Homeless expected its medical respite program to continue after the cooperative agreement, through payments from Medicaid payers. The five implementing sites had various payment approaches and they secured additional funding through other sources. The sites planned to continue the main program components, in some cases by maintaining the new staff or, more commonly, embedding the new functions into other staff workflows and EMRs. However, some sites thought they might need to cut back on services and functions to control their costs, or make slight modifications to the program. The awardee and participating sites demonstrated interest in scaling and replicating the program. One site added a new program in Year 3.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, National Health Care for the Homeless had begun actively pursuing strategies to sustain its program. All five implementing sites were preparing to sustain all major components of the program with some modifications by using new or existing funding streams from payers and other sources. Some sites reported discontinuing the exit survey because they felt it created

unnecessary confusion for participants. Some sites also reported shifting the focus of self-management goal setting; in one case to address any type of goal and in another case to focus on the first goal being health related. Other sites planned to modify the program to reduce its cost—for example, by selecting only certain services from the standardized model to continue. At the time, the awardee did not plan to sustain the data collection tool to track and report services delivered and it expected to end the data manager staff position. National Health Care for the Homeless had no plans to scale or replicate the program, although one of the overall goals of the intervention was to show up-and-coming respite programs a model for how to be successful.

C. Implementing the SSR plan: progress and changes

Sustainability. By Year 3, the awardee expected that payments from Medicaid payers, coupled with other funding sources, would allow implementing sites to sustain the program after the cooperative agreement ended. The five sites had various payment approaches, depending upon the program services offered, whether they were affiliated with an FQHC, and their state's Medicaid policies. To better cover costs, they also pursued several largely successful strategies to obtain additional funds from grants, foundations, and partnering hospitals. Hospitals were covering a significant share of the respite care costs. However, respondents expressed concerns about potentially losing funding from partnering hospitals as the programs secured relationships with payers.

Overall, the sites planned to continue the main program components, in some cases by maintaining the new staff or, more commonly, embedding the new functions into other workflows and EMRs. Most sites expected to keep their data analyst staff and data tracking functions because they found them useful to their broader quality improvement efforts. Some sites planned to use their data analyst in additional ways beyond the medical respite program. For example, one site was broadening the analyst's role to serve all of its programs, in part because many of the data requirements for the cooperative agreement were the same as public health reporting requirements. Another site was going to integrate health goal setting, prevention efforts (smoking cessation and influenza vaccination), and transition planning into its general standards of care. The partnering hospital for this site also embedded the patient confidence ratings into existing recuperation plans in the EMR. Another site was interested in continuing to track tobacco cessation and flu shot discussions because they also served as measures of patient engagement and staffing productivity.

Some sites planned to scale back or modify some services and functions. For example, some programs could return to a more acute care focus, with less emphasis on preventive care, namely influenza vaccination and tobacco cessation activities. All sites planned to either modify the survey questions or omit them altogether, which was not surprising because most staffed had reported that the surveys were confusing.

Scalability. The awardee reported that most sites wanted to grow their programs after the cooperative agreement ended. One site added a new medical respite care program, which was mostly funded (80 percent) from three local hospitals, with the remaining funding coming from insurance plans. In addition, the sites reportedly thought that some of the measures they had been collecting and reporting on for the program would be useful for the program on a larger scale.

Replicability. The program had not been replicated, although program staff attempted to share lessons learned with other implementing sites. At conferences, the awardee's executive director presented internal analyses showing an inverse relationship between Medicaid payments and hospital payments to respite programs across states—that is, as more Medicaid funding goes to medical respite services (for example, when a state expands the Medicaid program under the Affordable Care Act), hospital funding to support these services declines and program costs are not covered.

D. Factors associated with progress toward implementing the SSR plan

Demonstrating the value and ongoing costs of the medical respite program to potential funders was challenging for the awardee and its implementing sites. In particular, the sites found that hospitals appreciate the program but do not want to help fund it, particularly if Medicaid was reimbursing the site for respite care. The hospitals think the sites are receiving adequate compensation for their services, even though Medicaid isn't covering the full cost. According to one respondent, the executive director's messaging about this had been "tremendously helpful" to other medical respite programs trying build support for a payment model. In addition, sites were working with their closest hospital partners to build a business case to share with more reluctant hospitals, in hopes of convincing them that investing in medical respite care might make sense for them financially.

> "The hospitals want to refer. They want to have those bed turnovers. It's just a matter of, are they going to pay to do it?"
>
> —Site respondent

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. However, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or the sustainability, scalability, or replicability plans. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Nebraska Medicine

April 18, 2018

Colene Byrne (RTI International)
Alison Banger (RTI International)
Kevin Shang (Mathematica Policy Research)
Jia Pu (Mathemetica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Nebraska Medicine received a six-month extension to serve its patients through February 28, 2018. Enrollment of new patients ended on October 31, 2017.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

Nebraska Medicine used funding from HCIA R2 to create and implement the Remote Interventions Improving Specialty Complex Care (RIISCC) program. The awardee's program goals included using health coaching and remote patient monitoring (RPM) to help patients with diabetes (age 19 or older, not pregnant, and residing in the targeted areas) better manage their own care. RIISCC program staff recruited patients with a diagnosis of diabetes who had been hospitalized for any reason, not necessarily related to diabetes. The program provided 90 days of RPM after participants were discharged from the hospital, a follow-up visit after the 90 days, and an additional nine months of health coaching. Partners included a community hospital that is part of Nebraska Medicine and three primary care clinics where participants went for their 90-day visits (Table I.1).

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	To provide RPM and health coaching to individuals with type 2 diabetes and high likelihood of hospital readmission who reside in targeted areas of the greater Omaha, NE, metropolitan area
Major innovation	RPM and health coaching to improve self-management of diabetes
Program components	Care management, telemedicine, patient and family engagement, home care
Target population	Residents of Douglas, Sarpy, and Cass counties (in the greater Omaha metropolitan area) who live within 30 miles of Nebraska Medicine (NM) main campus, are age 19 or older, with diagnosis of type 2 diabetes, and were recently discharged from the hospital
Theory of change/ theory of action	By providing early and timely post-discharge health coaching and RPM and incentives for patients to participate, patients will better manage their diabetes and care at transitions, which will result in: (1) reduced morbidity and mortality, (2) avoided hospital and emergency admissions related to diabetes, and (3) reductions in total costs of care for the target population
Payment model	New fee-for-service (FFS) payment with shared savings and bundled per-episode payment for telehealth services
Award amount	$9,993,626
Effective launch date[a]	December 22, 2014
Program setting	• Academic medical center • Community health care clinics • Participants' homes
Market area	Urban
Market location	Douglas County, NE, where the city of Omaha is located, and Sarpy and Cass counties, both of which are south of Douglas County
Target outcomes	• Improved blood pressure and hemoglobin A1c • Fewer all-cause unplanned readmissions hospital-wide • Fewer ED visits • Reduction in body mass index • Reduction in morbidity and mortality due to diabetes and obesity • Increase in the percentage of participants receiving diabetes eye and foot exams

[a]After the initial planning period, the awardee's program became operational as of this date.

ED = emergency department; RPM = remote patient monitoring.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, the awardee partially achieved enrollment effectiveness, having enrolled 1,903 participants—76 percent of its enrollment target—by the end of the initial cooperative agreement. Second, the awardee effectively delivered intervention services. It provided consistent services to participants, engaged local providers, achieved high levels of participant satisfaction, and, based on its clinical performance metrics, helped participants control their diabetes. The program provided an additional level of support for participants, enabling physicians to focus on direct patient care. Third, the awardee ensured it had the staff and training to deliver the services. Although it encountered some challenges with staff retention, it improved its approach to staffing and operations over the cooperative agreement. The health coaches and medical assistants received ongoing education from biweekly meetings with the program's medical director and the diabetes center clinical director, working through examples of caring for program participants. Fourth, the awardee successfully engaged participants, as demonstrated by high levels of participant activation and satisfaction. Finally, implementation staff reported the program had a positive effect on the delivery of care.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Nebraska Medicine's RIISCC program, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants and the results of our efforts to identify a matched comparison group.

Payment model. Nebraska Medicine continued developing a bundled payment model. In this model, the payer reimburses the Nebraska Medicine program for a set amount per patient for one episode of telehealth services (for example, 90-day RIISCC intervention) provided by the whole RPM team, including primary care physicians (PCPs), nurses (leads and health coaches), medical assistants (MAs), dietitians, ophthalmologists, and other staff, rather than paying for each individual service. The awardee describes the approach as consistent with Model 3 in the Bundled Payments for Care Improvement initiative, in which Medicare continues to make FFS payments and reconciles the payment later based on a bundled payment amount determined by CMS. Based on the difference between the aggregate expenditures and the target price, Medicare then makes a payment or recoupment. Savings would occur if the telehealth services improve participants' outcomes and reduce their use of inpatient and ED care.

To determine a reasonable charge rate for telemedicine services provided to patients with diabetes, the awardee conducted a preliminary analysis to estimate the cost of providing the telemedicine services to a participant during the 90-day RIISCC intervention, using Nebraska Medicine's internal claims data. The awardee plans to conduct payer-specific analyses using claims data recently received from Medicare and Nebraska Blue Cross and Blue Shield.

Sustainability plans. Nebraska Medicine focused on engaging organizational leaders as its sustainability strategy in the third program year. The awardee began outlining the program's

benefits relative to its cost, and make plans to meet with organizational leaders and payers to discuss funding and sustaining the program.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of non-clinician staff on their perceptions of the effect of RIISCC on the delivery of care. The non-clinician staff survey was fielded around the start of the third program year with a sample of 23 potential respondents and achieved a response rate of 74 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that Nebraska Medicine was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

RIISCC program staff originally recruited patients hospitalized at Nebraska Medicine who had a diagnosis of type 2 diabetes and who met eligibility criteria (that is, age 19 or older, not pregnant, English speaking, and residing in the targeted geographic area.) Patients did not have to be hospitalized for a reason related to their diabetes; same-day procedures (such as a colonoscopy) or other reasons for admission (such as a motor vehicle accident) qualified patients for recruitment. A lead nurse briefly introduced the program to patients in the hospital with take-home information. After discharge, the program sent the same information to the patient's home, and health coaches, who are registered nurses (RNs), called to recruit patients. Nebraska Medicine originally intended for all participants to take the RPM equipment home directly from

the hospital but quickly realized that some participants required home delivery and installation, so it added this option.

Nebraska Medicine refined recruitment and enrollment processes as it learned from its implementation experience. Primary care medical teams and diabetes educators in both the inpatient and outpatient areas were sources of referrals during patient hospitalizations and initial follow-up visits post-discharge. Patients also received invitations to participate via messages sent through Nebraska Medicine's patient portal. Beginning in the second program year, if the patient agreed to participate, the nurse offered the patient two options: (1) an MA comes to the participant's home to deliver, install, and demonstrate use of the RPM equipment, or (2) the participant comes to Nebraska Medicine's telemedicine hub to learn how to use the equipment before taking it home.

Recruitment was a key challenge for Nebraska Medicine. Its initial enrollment targets proved unrealistic as the team began enrolling patients, and the number of patients identified as eligible for the program was lower than projected. The acceptance rate of patients who were approached to enroll (approximately 50 percent) also was lower than anticipated. A number of programmatic changes increased enrollment in the first program year, including adding Nebraska Medicine-Bellevue hospital (part of Nebraska Medicine system) to the program as another source of patients. A clinician with extensive electronic medical record (EMR) expertise helped the team refine the criteria for identifying patients in the EMR system, which doubled the number of eligible patients. In the second program year, awardee program staff made additional changes to their recruitment strategy. Experts in recruitment and enrollment for research studies at the University of Nebraska Medical Center advised the team on how to improve the protocols and scripting to increase the program acceptance rate. Other recruitment improvements included the following:

- Expanding the recruitment window (from three days following discharge to four weeks)
- Recruiting patients who had previously declined after an earlier hospitalization
- Discontinuing in-hospital recruitment (staff found that patients were already overwhelmed at the time of discharge and that hospital recruitment was not productive)
- Increasing the number of recruitment calls made to patients
- Offering participants the option of picking up the RPM equipment or having it delivered
- Reviewing the recruitment scripts with experts in recruitment and patient engagement
- Developing a cohesive set of marketing materials that branded the program consistently

In the third program year, program staff made two changes: (1) the program was offered to patients who speak Spanish (previously only offered to English speaking patients), and (2) eligibility expanded to Nebraska Medicine patients with diabetes who were hospitalized within the past year at other hospitals besides the two participating hospitals.

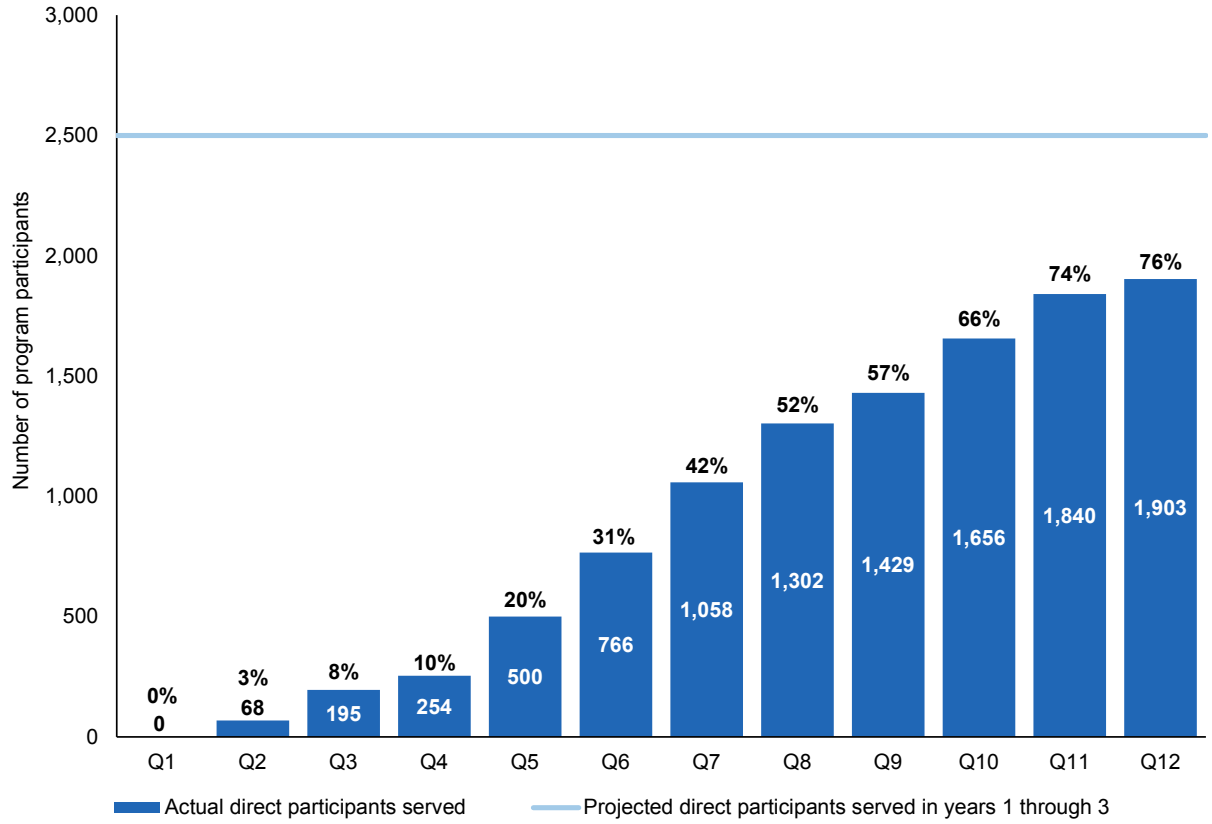
b. Evidence of enrollment effectiveness

Overall, Nebraska Medicine partially achieved enrollment effectiveness. In November 2016, CMS approved the lowering of Nebraska Medicine's three-year direct projected enrollment from

3,300 to 2,500 participants. This change occurred primarily because the number of patients identified as eligible for the program was lower than projected, and the acceptance rate of patients who were approached to enroll (approximately 50 percent) was lower than anticipated. In addition, Nebraska Medicine delayed enrollment early in Year 1 when the principal investigator, along with other staff, resigned, and new staff had to be hired. The awardee enrolled 1,903 participants from December 2014 (when it launched its program) through August 2017, which represents about 76 percent of its revised 2,500 three-year projected participants (Figure II.1). When measured against the three-year projection set at the end of Year 1 (enough time for the awardee to gain a better understanding of the number of eligible participants but before the awardee began planning for the end of the cooperative agreement), the awardee met 58 percent of its projection.

By the end of the second program year, Nebraska Medicine had improved enrollment and was close to or met its monthly and quarterly enrollment targets. However, for almost two months in the last program quarter, Nebraska Medicine suspended recruitment and enrollments due to delays in waiting to hear whether awardees received no-cost extensions. This resulted in uncertainty as to whether Nebraska Medicine could offer new patients the full 90 days of the RPM intervention and nine months of health coaching. Its percentage of its goal met would have been higher had there not been two months of no enrollments.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. In November 2016, CMS approved the lowering of Nebraska Medicine's three-year direct projected enrollment from 3,300 to 2,500 participants.

c. Barriers and facilitators associated with enrollment effectiveness

Nebraska Medicine's progress in meeting its three-year enrollment goals was influenced by several barriers and facilitators.

Barriers included the period of uncertainty as to when the program funding was ending, as noted above; issues with correctly identifying eligible patients; and staffing issues. A barrier that impeded identifying eligible patients was the knowledge and training of the staff who were generating the lists of eligible patients through their EMR. For several months, the list did not include patients who met the program's eligibility criteria because the algorithm was incorrect—either due to staff's lack of understanding the EMR data, or because untrained staff ran the reports incorrectly. Another barrier was the recruitment and turnover of the MAs, who deliver and set up the RPM equipment at participants' homes. At times, a shortage of MAs caused delays in starting services for recruited participants.

Correcting and improving the algorithm for generating patient lists from the EMR substantially increased the pool of eligible patients. Despite the other efforts to market the program and improve patient acceptance, however, patient interest in participation remained a barrier. The percentage of eligible patients who agreed to participate remained around 50 percent over the three-year program; however, recruitment did improve because of the larger pool of eligibles. The surveyed staff felt that the time commitment and lack of understanding the program were two main reasons patients chose not to participate (82 percent and 47 percent, respectively).

In addition to steadily improving patient recruitment and enrollment as described above, Nebraska Medicine used several strategies to increase internal awareness and support of the program. It developed an educational video about the RIISCC program and posted it on Nebraska Medicine's intranet. In addition, it leveraged the medical director as a physician champion. Getting buy-in from patients' primary care medical teams, diabetes educators, and staff on the inpatient unit also helped enrollment. For example, hospitalists do not typically have the established relationships that primary care and other physicians have with eligible potential participants, so program staff worked with the hospitalists to encourage their patients to participate.

Another facilitator was technical assistance from the implementation and monitoring contractor. RIISCC program staff implemented several of the contractor's recommendations. For example, informational packets about the program were updated to include photos of the health coaches and MAs so that participants could put a face to the name of staff with whom they would be working. RIISCC staff also worked to enhance intentionality with participants by asking them to think about logistical issues such as where they would put the RPM equipment and who they would contact if they needed help with the equipment.

RIISCC staff also reported that gift cards were effective participation incentives. Participants received a $10 gift card as an incentive at certain stages of the program (enrolling in the program, first data upload, 30-day data upload, and 90-day visit). They received an additional $10 gift card for returning the RPM equipment, totaling $50 in gift card incentives.

2. Delivery of program services

a. Description of and changes to service delivery model

Nebraska Medicine used funding from HCIA R2 to create and implement the RIISCC program. The goal of the program was to use health coaching and RPM to help patients with diabetes better manage their own care. The RIISCC program provided 90 days of RPM, a follow-up clinic visit after the 90 days, and then an additional nine months of health coaching.

Nebraska Medicine's partners included Nebraska Medicine-Bellevue, a community hospital that is part of Nebraska Medicine, and three primary care clinics where participants went for their 90-day visits. The awardee's program goals were to (1) improve participants' management of diabetes through better blood glucose and blood pressure control; (2) improve diabetes-related care following hospital discharge, including receipt of diabetes eye and foot exams; (2) reduce all-cause admissions and ED visits within 90 days of discharge; (3) reduce hospitalizations for uncontrolled diabetes; and (4) reduce the total costs of care for participants.

> "We'll make sure you [patient in program] understand ways of managing your diabetes better, that you understand what your blood glucose numbers mean, you understand how your medication works, and what you need to do when things aren't working well for you. It's the problem-solving."
>
> —Certified diabetes educator

The RIISCC program recruited hospitalized patients diagnosed with type 2 diabetes who met eligibility criteria (that is, age 19 or older, not pregnant, English or Spanish speaking, and residing in the targeted areas.) Patients could be hospitalized for any reason, including a same-day procedure; the hospitalization did not have to be related to their diabetes. Health coaches recruited eligible patients within four weeks of discharge. Initially, Nebraska Medicine recruited patients in the hospital or, if they were unable to contact the patient before discharge, via telephone with the MA delivering the equipment. However, Nebraska Medicine soon realized that hospital recruitment was ineffective and dropped this portion of its recruitment effort. In parallel, the organization built a new telemedicine hub where the program's offices relocated and where patients and physicians could test RPM equipment. The RPM equipment transmitted information about the participant's weight, body mass index, blood pressure, and blood glucose values to the health coach and providers.

Participants received weekly calls from their health coach to discuss progress and challenges toward health goals. After 90 days of RPM, participants visited one of the primary care clinics to return their RPM equipment, check their hemoglobin A1c level, receive a foot and eye exam, and receive nutritional counseling. After the 90 days, participants received nine months of monthly coaching calls where they could continue to discuss their health concerns and goals with their health coach and develop strategies to meet those goals.

The RIISCC program is based on the theory of change/theory of action that health coaching, as delivered in the program, can improve patient self-management of diabetes (for example, better blood glucose control, lower blood pressure), which will improve health outcomes (for example, morbidity, mortality) and reduce hospitalizations related to diabetes and their associated costs.

b. Evidence of service delivery effectiveness

Overall, Nebraska Medicine was effective in implementing its program. Although it refined the program over time, the core services (RPM, 90-day visit, and health coaching) remained consistent throughout the duration of the cooperative agreement. Nebraska Medicine recovered quickly from losing its original principal investigator and several staff after program implementation. It overcame staffing challenges and hired, trained, and retained qualified and highly committed staff. Patient engagement and satisfaction with the program were high for participants who completed the program. Service delivery effectiveness was driven by the close relationships the health coaches formed with their patients, effective integration of the RIISCC program with the larger Nebraska Medicine organization, and the awardee's approach to continuous program improvement.

Delivery of intervention services. According to program leaders and staff, the RIISCC program was very successful in delivering RPM and health coaching services to patients with diabetes. Eighty one percent of the staff completing the non-clinician survey early in the third program year felt that the program had a positive impact both on the quality and efficiency of care or services provided to participants. Eighty-two percent felt that the program resulted in better treatment plans and/or outcomes among participants. Eighty-eight percent felt that the RIISCC program had a positive impact on access to care or services for participants.

Staffing and training. After overcoming some staffing problems in the first program year, Nebraska Medicine was partly successful in recruiting, hiring, and retaining the program staff to deliver services. Nebraska Medicine had ongoing challenges filling the MA roles. To address this, it maintained a continuous job posting for MAs and worked with Nebraska Medicine's human resources department to optimize recruiting, hiring, and orienting new hires to the organization.

In addition, uncertainty about the program's end date resulted in several staff leaving for other positions in the middle of the third program year. Once the awardee learned it had received a six-month no cost extension, it identified internal staff to shore up the team (for example, handle administrative and clerical work) and ensure that the remaining clinical staff could deliver patient care.

Nebraska Medicine reported that as of August 2017, it had exceeded its hiring goal and had achieved an 86 percent retention rate for staff. In addition, Nebraska Medicine achieved its three-year training target. Sixty-five percent of the surveyed staff had received formal training from the program and almost all reported receiving informal training, most commonly technical assistance (81 percent), help from a colleague (94 percent), individual supervision (76 percent), mentoring (75 percent), and shadowing other staff (71 percent).

Recruitment and engagement of providers Partnering with the physicians was an important aspect of the program and generally worked well, although challenges remained. Twenty-five percent of the surveyed non-clinician staff felt that difficulty working across collaborating organizations was a barrier to effective implementation.

Interviewed program staff noted a growing number of local physician champions who were familiar with the program and appreciated the services it offered to the participants. Physicians found the RPM data beneficial because it provided a reliable view of the participant's health over time. One blood pressure or blood glucose reading in an office might be unreliable, but daily measurements taken in a consistent way offered a clearer view into the participant's health and any potential changes to their care that might be necessary.

Engagement of program participants. According to program staff and leaders, Nebraska Medicine was partly successful in engaging program participants. Forty-one percent of the surveyed staff strongly agreed and 24 percent somewhat agreed that the program has successfully engaged participants. Seventy-one percent felt that participants' engagement in the program was one of the most helpful factors for achieving program goals. Nebraska Medicine also conducted several surveys and analyses to better understand participant engagement, including participant satisfaction, participant activation, and patterns in enrollment and disenrollment. The scope of this evaluation does not include an assessment of the methodology and analytic approach used by Nebraska Medicine; however, the results of these additional analyses are likely impacting program staff's perception of the program. See Section C, Assessment of perceived program effects on the delivery of care and outcomes, for a discussion of these findings.

c. Barriers and facilitators associated with service delivery effectiveness

Nebraska Medicine's ability to effectively deliver intervention services was influenced by several barriers and facilitators as discussed below. Barriers included recruiting and retaining MAs and other staff; difficulty improving patient participation and disenrollments; and the need to repair or replace the eye exam equipment used at the 90-day clinic visits. Facilitators included expanding eligibility criteria, delivering RPM as planned with close adherence to policies and protocols, using health information technology to support service delivery, continuous program monitoring and improvements by the RIISCC team, integrating program and services within the Nebraska Medicine system, participation by clinics outside the Nebraska Medicine system that serve patients in the program, and strong patient engagement and rapport with the health coaches.

Recruiting, hiring, and retaining MAs were ongoing challenges for the program. The MA role in RIISCC was unlike traditional MA roles in a hospital or physician's office; it required visiting homes to deliver, install, and demonstrate the RPM equipment; performing tasks that were more clinically complex (for example, taking eye screening photos); and engaging more with participants. Nebraska Medicine maintained a continuous job posting for MAs and worked with the organization's human resources department to ensure it fully understood Nebraska Medicine's hiring processes (for example, when new employee orientation takes place, deadlines for returning completed paperwork). Forty-four percent of the surveyed non-clinician staff reported that staff turnover and unfilled positions posed a major barrier to achieving program goals.

Another barrier was the need to more effectively close the communication loop between patients' PCPs and health coaches. The program had alerts for high/low blood pressure and blood glucose levels. The health coaches could send an alert via the EMR (in the same way that an aberrant lab result would be flagged in the record) and would call the physician if a critical result indicated the patient had to be seen immediately. However, health coaches were sometimes uncertain whether the physicians saw their notes and recommendations. Communication was simpler when dealing with physicians within the Nebraska Medicine system because of the shared EMR, which indicates when a physician has read a note, but the coaches and certified diabetes educators were often not sure what information physicians received.

Another related challenge was that the physician of record did not necessarily have a close or recent relationship with the participant, or the physician was not aware that his or her patient was in the program. For example, a physician would start receiving data on and notifications about a patient's blood sugar levels but had not seen the patient for an office visit nor spoken to the patient in more than a year. Physicians were sometimes frustrated because they did not want to be responsible for or receive data until they had a visit with the patient.

The hiring of a telehealth clinic manager was key for streamlining operations and connecting the RIISCC program with other parts of the Nebraska Medicine organization. Separately from the RIISCC program, Nebraska Medicine created a new position of telehealth clinic manager for the Nebraska Medicine organization and filled it with an RN who worked under the new telehealth director. Both worked to improve and advance Nebraska Medicine's broader telehealth program as well as the RIISCC program, with the mission, according to program managers, "to spread the program's efforts to other realms within Nebraska Medicine". They built relationships across Nebraska Medicine and increased awareness among Nebraska Medicine physicians. The telehealth clinic manager focused on integrating telehealth into the existing clinical workflow, which had been lacking. To do so, she used her knowledge of clinical personnel, equipment managing and processing, policies and procedures, and the other day-to-day concerns of the clinical staff. The telehealth clinic manager also helped meet staffing needs by working through the Nebraska Medicine hiring bureaucracy. Further, she worked to improve collaboration, communication, and coordination among Nebraska Medicine clinical staff who treated the same patients to create a more seamless experience for participants and prevent the duplication of clinical efforts.

Service delivery was also facilitated by biweekly training sessions led by the RIISCC medical director and director of Nebraska Medicine's Diabetes Center—both are also physician champions. These sessions took place with the lead nurses, nurse coaches, MAs, and other clinical program staff. The team also shared promising practices at short meetings that took place every three weeks. An internal staff website was established to keep information updated and accessible to all staff, even if they did not attend the meetings.

The MAs and health coaches adhered to protocols for interacting with participants while providing care. The nurse health coaches spoke with the participants regularly, allowing them to develop relationships with the participants and better understand their health and health needs. The health coaches could observe and follow up on potential changes that might indicate a need for participants to see their physician soon or that could indicate a potential emergency. The health coaches also had the time and capacity to delve into the participants' motivation and

challenges and work with them to find solutions. The program was beneficial because it enabled staff to intervene with participants who have higher medical needs that might not necessarily require a doctor's visit but do require more intensive monitoring and follow-up. Multiple program leaders and frontline staff mentioned that the strong rapport between the nurses and participants was a key reason for program success. In some instances, the lead nurses and health coaches were more up to date on diabetes treatment protocols and best practices than were the PCPs with whom they were working. Because of their regular meetings and trainings, the lead nurses and health coaches were abreast of new treatment guidelines and recommendations, and, with the assistance of the physician leads, implemented them into their coaching. However, health coaches could feel uncomfortable stating or implying that a physician was not using the most current information to treat his or her patients. The physician champion and health coaches worked together to brainstorm ways to discuss more appropriate goals for the participant with the PCP and ensure that all participants were being treated according to the most up-to-date evidence. The physician champions noted that their role supporting the health coaches evolved as they gained more confidence and skills in diabetes management.

Program leaders identified providing training, supervision, and support for program staff and fostering communication and collaboration among staff as major facilitators to delivering program services. The project leadership team was attentive to detail and the need to ensure proper staff training. In late 2016, project leaders realized they were not meeting their monthly recruitment targets. After investigating, they realized that not all health coaches were fully trained in how to extract and review the list of potentially eligible patients. They subsequently cross-trained all the coaches and ensured they could all complete the task.

> "It really comes down to patients and their willingness to take care of themselves. That's why we fail in clinics, because we have to rely on the patients' compliance to do these things. If you see them every three months or every six months, they don't. I have a diabetic [patient] who drinks regular soda every day and I see him every three months and he just can't stop. So, if he were to get called every day, he'd make a difference."
>
> —RIISCC medical director and physician champion

Patient engagement was seen as the critical component for the program's success and a driver for helping patients improve outcomes, such as improved A1c levels and medication adherence. Program staff described a wide array of approaches to working with participants aimed at meeting the participant where he or she is. For example, a CDE noted that a goal to walk 10,000 steps a day might be feasible for a younger person with a 9-to-5 job and without significant family responsibilities, but might simply be discouraging for someone who works two jobs and has caregiving responsibilities at home. The health coaches had consistent contact with participants, and took the time to figure out what was driving their behavior. This allowed the coaches to more readily identify behavior changes and approaches to address challenges the participants were facing in managing their diabetes.

Nebraska Medicine also implemented information technology solutions that support the flow of information from the program to participants' PCPs. It developed an interface enabling patient data from the RPM system to transfer into the EMR system at Nebraska Medicine and at Nebraska Medicine-Bellevue hospital. The interface facilitated communication and collaboration between the health coaches and participants' PCPs. A new section of the EMR system was set aside for telemedicine encounters; it displayed the RPM data along with the coaching notes.

Clinicians involved in the program were highly satisfied with the interface, which supported their clinical workflows. Nebraska Medicine's charting improved significantly over the three years. Nebraska Medicine implemented more structure and added smart phrases to its EMR, which enabled the health coaches to chart more quickly.

Nebraska Medicine also compiled and electronically linked not only the data available from the EMR and RPM systems but the various types of program data that staff collected as well. This integration across its information systems ensured a more comprehensive view of a participant's progress and care. For example, during enrollment, staff collected and recorded data on demographics, including contact information. Information gathering could occur in the home environment by the MAs while setting up equipment, or as part of the baseline Patient Activation Measure (PAM) survey, which was used to measures participants' ability to take care of themselves and make confident decisions about their health, and was a source of data for a self-monitoring measure. The first patient uploads of RPM data—baseline weight, blood pressure, and hemoglobin A1c levels—were also collected during the home visit. Data on health care utilization were sometimes collected during calls with the participants. Foot exam and hemoglobin A1c test results, follow-up PAM data, and participant satisfaction data were collected at the 90-day visits. Some data, such as those regarding hospital admissions or ED visits, were also collected post-intervention during the follow-up calls. RIISCC program staff used tablet computers and REDCap (Research Electronic Data Capture, a web-based software tool) to capture and upload data from the participants and manage the program database. The streamlined process and the new unified database helped the awardee to evaluate health outcomes and conduct the economic analyses necessary for developing a payment model.

C. Assessment of perceived program effects on the delivery of care and outcomes

> "[The health coaches] are able to have a deeper relationship with the patient—they have time and they have these 90 days to get to know somebody. The physicians don't often get to the root of why [their patients] are not doing what they know they should be doing. This is what this program allows: you get to take time and tease out that information over time."
>
> —RIISCC physician champion

In interviews, program leaders and staff were uniformly positive about the program itself and the outcomes they saw. Health coaches reported seeing behavior change in their patients (increased exercise, improved diet); the CDE reported receiving very positive feedback from participants at the 90-day visits; and Nebraska Medicine documented improved participant outcomes, including lower A1c levels and increased measures of participant activation. Although not all patients engaged with the program, those who did so were usually successful in achieving improved outcomes. The health coaches also observed that participating in the program encouraged patients to re-engage with their PC.

This assessment is also clearly reflected in the responses to the survey conducted at the beginning of the third program year, which showed a strong belief in the positive impact of the program. Table II.2 provides an overview of these findings. Most of the non-clinician staff surveyed indicated that the program had a positive impact on the quality of care and services provided to participants. Of the staff who felt the program was effective in achieving its goals, more than half attributed this to better participant or patient treatment plans and outcomes. The majority of surveyed staff believed that the program had positive impacts on the delivery of care and outcomes such as participant satisfaction and quality of life. In addition, most staff indicated

the program was very or somewhat effective in achieving its goals. Eighty-seven percent felt the program was worth the effort.

Table II.2. Staff perceptions of RIISCC program's effects on care

The RIISCC had a positive impact on:	Percentage
The quality of care and services you provide to participants	84
Your ability to respond in a timely way to participant needs	75
Your ability to provide care or services that are responsive to participant preferences, needs, and values	81
Access to care or services for all participants	88
Achievement of participants' health goals	94
Participant satisfaction	100
Participant quality of life	94
Care coordination	88

Source: HCIA R2 evaluation survey of awardee's non-clinician staff, fall 2016. The survey had 16 respondents overall, but for some survey items, one or two responses were missing.

RIISCC = Remote Interventions Improving Specialty Complex Care.

Nebraska Medicine regularly assessed its program effectiveness internally. Evaluating the data and methods used in these analyses is beyond the scope of this evaluation, but likely impacted program staff's perceptions of the program. Specifically, Nebraska Medicine conducted patient satisfaction and patient activation surveys and analyzed enrollment and disenrollment trends.

Nebraska Medicine noted consistently high levels of satisfaction throughout the program for participants who complete the program. Ninety-five percent of participants in its most recent survey agreed or strongly agreed that their overall experience with the RIISCC program has been positive; the same percentage said they would recommend the program to a friend. The results of the PAM survey were also strong and mean patient activation scores increased significantly from baseline to the end of the 90-day intervention. Finally, Nebraska Medicine extensively analyzed its enrollments and disenrollments to better understand the factors that might be influencing patients to enroll in and remain in the program. The findings suggest the program might not have been as successful in engaging its higher-risk patients.

Nebraska Medicine staff's perceptions of the program effects were also likely based on the awardee's internal analyses of the impact of services on outcomes, including hospital readmissions, diabetes control as measured by improvement in hemoglobin A1C tests and surveys measuring patients' ability to self-manage their diabetes with confidence, and improved understanding of the disease process. Nebraska Medicine's analyses indicated that some demographic groups and higher-risk patients (for example, those with higher blood pressure and/or A1c levels, smokers) are also more likely to disenroll. In its analyses, Nebraska Medicine measured financial impact by readmission data and total cost-of-care data. Its impact results have reportedly been very favorable.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. It is important to note that our assumptions are predicated on the awardee's theory of action being accurate in terms of the planned processes resulting in the desired outcomes, which we did not evaluate.

The main effects of this program are likely to be a reduction in ED visits, hospitalizations, and readmissions, leading to a decrease in total cost of care. By the proposed theory of action, there would also be an expected rise in outpatient utilization. If participants effectively manage their diabetes and receive regular outpatient care, it is reasonable to expect a change in targeted outcomes within two years. The delay in implementation in Year 1 was temporary and short and had a small impact on implementation effectiveness. However, Nebraska Medicine reported difficulties with enrollment, disenrollment, and patient engagement, particularly for the sickest and highest-risk patients. Given that highest-risk patients are less likely to enroll in the program and more likely to disenroll, meeting the targeted outcomes might be challenging, as these patients would have more potential to improve their management of diabetes and avoid hospitalizations and ED visits. Nebraska Medicine plans to continue the program based on clinician referral of high-risk patients (for example, poor blood glucose control) rather than targeting only recently hospitalized patients. Patients' hospitalizations would be one criterion clinicians could use to determine whether the patient was at high risk. A change to clinician referral of patients could result in awardee staff more effectively meeting the program's targeted outcomes by reaching patients who are at higher risk of hospitalizations and morbidity and mortality, and who might be more amenable to the intervention.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Nebraska Medicine's RIISCC program, baseline characteristics of the treatment group, and the results of our efforts to identify a matched comparison group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: Nebraska Medicine

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	482[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	149[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,649
Likelihood of all-cause hospitalizations	553
MDE sample size requirement to detect 20% effect	
Total expenditures	412
Likelihood of all-cause hospitalizations	138
Participation/Selection bias of concern	Yes, patient self-selection high/high refusal rate
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	Some issues, but probably surmountable. Expect to select a comparison group.
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We plan to conduct a rigorous evaluation of Nebraska Medicine's RIISCC program by using a comparison group of Medicare beneficiaries discharged from other hospitals in Omaha and by matching on demographic characteristics, prior use, and secondary diagnoses other than diabetes. The matching process is complete and has resulted in a comparison group that closely resembles the treatment group in terms of demographic characteristics, dual enrollment in Medicare and Medicaid, hierarchical condition category (HCC) score, and prior Medicare expenditures.

B. Characteristics of Medicare and Medicaid participants at baseline

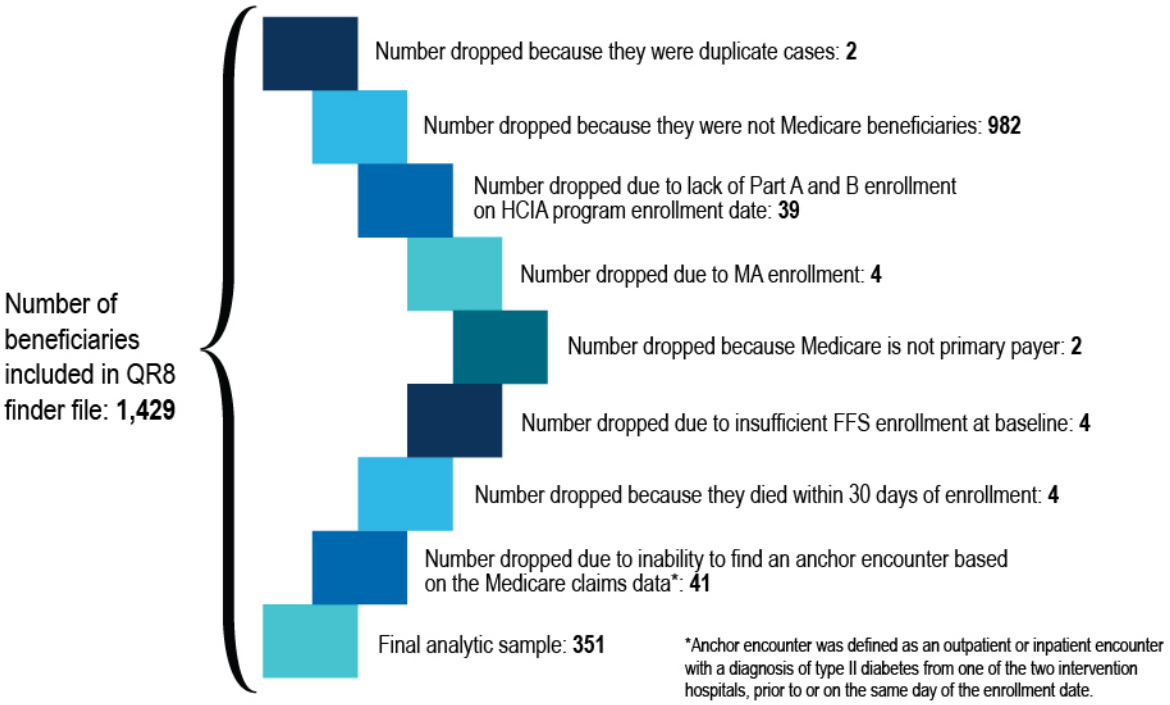
This section summarizes baseline characteristics of the treatment group, which were measured during the 12 months before each beneficiary's enrollment date in the RIISCC program. The treatment group consists of beneficiaries admitted to an inpatient or outpatient stay at Nebraska Medicine with a primary or secondary diagnosis of type 2 diabetes and who accepted the invitation to participate in the program.

Nebraska Medicine began to enroll Medicare beneficiaries in the RIISCC program in January 2015. As of the end of October 2017, the awardee had enrolled 2,003 beneficiaries.[4]

In presenting baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when they became eligible for awardee-provided services (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before enrollment). In addition, they must have been enrolled in the program on or before November 30, 2016, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. Furthermore, they must have had an inpatient or outpatient encounter with a primary or secondary diagnosis of type 2 diabetes at one of the two intervention hospitals within one year before enrollment. The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. After we excluded beneficiaries who did not meet the above criteria, 351 participants were included in the analysis of baseline characteristics for this report.

[4] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

Figure III.1 Reasons for exclusion of program enrollees from impact analyses



Participants were more likely than the Medicare population as a whole to be black, originally entitled to Medicare because of a disability, and enrolled in both Medicare and Medicaid (Table III.2). The differences in the populations for the latter two categories were marked. A Mathematica analysis of a random 5 percent sample of Medicare beneficiaries found that 24 percent of beneficiaries nationally were originally entitled to Medicare because of a disability. Among Nebraska Medicine's program participants, the figure was 45 percent. Furthermore, among program participants, 30 percent of Medicare beneficiaries were also enrolled in Medicaid, compared to 20 percent nationally.[5]

Program participants by definition have a least one chronic medical condition—that is, diabetes. Moreover, all participants, again by definition, were discharged at least once for inpatient or outpatient care in the baseline year. It is therefore not surprising that their Medicare utilization and expenditures are well above average. The mean expenditure per beneficiary per month (PBPM) in the baseline year was $1,745—more than double the national average of $792 (Table III.3). Acute hospital admissions, at 952 per 1,000 beneficiaries per year, were far greater than the national average of 274 per 1,000 beneficiaries in 2014. ED visits and observation stays for the treatment group were also well above national averages.[6] In addition, there was an

[5] See the report on Medicaid enrollment from the Medicare Payment Advisory Commission and the Medicaid and CHIP Payment and Access Commission at http://www.medpac.gov/-blog-/medpacblog/2015/01/09/new-data-book-on-beneficiaries-dually-eligible-for-medicare-and-medicaid

[6] For national average rates, see CMS's "Public Use File; New Data on Geographic Variation," available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html.

upward trend in expenditures and utilization in the last quarter before enrollment, corresponding to the high probability of inpatient or outpatient care. The total average PBPM Medicare payment rose from $1,369 in the first quarter of the baseline year to $3,216 in the last quarter before enrollment (Table III.3). This trend has important implications for selecting a comparable cohort of Medicare FFS beneficiaries as the comparison group.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Nebraska Medicine's program through November 30, 2016

	All participants (N = 351)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	112	32
65 to 74	172	49
75 to 84	55	16
85 and older	12	3
Gender		
Female	206	59
Male	145	41
Race		
White	263	75
Black	81	23
American Indian, Alaska Native, Asian/Pacific Island American, or other	2	0.57
Hispanic	3	0.85
Original reason for Medicare eligibility		
Old age and survivor's insurance	185	53
Disability insurance benefits	157	45
End-stage renal disease (ESRD)[a]	9	3
Hospice[b]		
Medicare/Medicaid dual status, percentage dual[b]	107	30
HCC score[c]		Statistic
Mean		1.96
25th percentile		1.06
Median		1.65
75th percentile		2.51

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined based on the enrollment date provided by the awardee. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

Table III.2 *(continued)*

[b]Identified in the last month of each beneficiary's baseline year.
[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.
CMS = Centers for Medicare & Medicaid Services; FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Nebraska Medicine's program through November 30, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of participants	351	324	336	351	351
Average Medicare expenditures PBPM[a]					
Total	1,745 (141)	1,369 (283)	1,108 (126)	1,194 (154)	3,216 (291)
Acute inpatient	833 (108)	586 (230)	396 (78)	466 (103)	1,822 (262)
Inpatient other[b]	15 (10)	33 (33)	19 (14)	5 (5)	6 (6)
Outpatient[c]	387 (35)	275 (40)	262 (39)	304 (47)	684 (68)
Physician services	361 (20)	297 (31)	294 (23)	296 (27)	544 (29)
Home health	68 (11)	61 (18)	61 (16)	63 (14)	87 (17)
Skilled nursing facility	47 (15)	87 (35)	42 (27)	28 (20)	33 (19)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	34 (4)	29 (4)	34 (5)	32 (5)	41 (5)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	952 (84)	621 (159)	559 (94)	539 (99)	1,996 (150)
Outpatient ED visits[e]	1,867 (236)	1,394 (275)	1,191 (217)	1,558 (253)	3,217 (340)
Primary care visits in any setting	9,575 (418)	8,665 (612)	8,481 (535)	7,815 (533)	13,050 (580)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Primary care visits in ambulatory settings	7,340 (309)	7,386 (481)	6,865 (376)	6,374 (402)	8,612 (402)
Specialist visits in any setting	15,278 (941)	14,442 (1,477)	13,608 (1,188)	13,286 (1,045)	19,472 (1,060)
Specialist visits in ambulatory settings	10,604 (697)	10,895 (938)	10,558 (928)	9,935 (754)	10,974 (703)
Measures of any health care utilization					
Percentage with a hospital admission[d]	53 (3)	12 (2)	11 (2)	10 (2)	40 (3)
Percentage with an outpatient ED visit[e]	66 (3)	18 (2)	16 (2)	20 (2)	50 (3)
Percentage with a 30-day readmission among all discharges	17 (3)	10 (5)	23 (6)	16 (5)	16 (6)
Percentage of participants with a readmission among all participants	6 (1)	1 (1)	3 (1)	2 (1)	1 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

C. Quality of matched comparison group

Nebraska Medicine is an academic medical center in Omaha, Nebraska. Staff from its RIISCC program enrolled patients with type 2 diabetes from medically underserved areas who had been recently discharged within the prior year from one of two program hospitals. The goal

of the program was to use health coaching and RPM to help patients with diabetes better manage their own care. Together with Nebraska Medicine-Bellevue, an affiliated community hospital, Nebraska Medicine launched its program in December 2014, partnering with three primary care clinics where participants received their 90-day follow-up visits as part of the program.

Because the intervention could not be implemented using an experimental design, we rely on a matched comparison group design to examine the impact of the RIISCC program. The results presented in this narrative demonstrate the high level of covariate balance achieved on a range of beneficiary characteristics that are likely to affect participation in the intervention and outcomes.

The goal of matching is to minimize the effect of nonrandom selection of individuals into the treatment group by constructing a matched comparisons group that appears similar to the treatment individuals on key observable characteristics that affect treatment status and outcomes. For the RIISCC intervention, key characteristics for matching include a baseline A1c testing indicator, as well as individual demographics, health status, and expenditure and utilization history. Limiting the comparison group to those Medicare beneficiaries whose observed characteristics closely match those of the treatment group will also reduce unobserved differences between the two groups if those characteristics are correlated with matching variables.

1. Description and identification of the treatment group

The first step of the RIISCC program was identifying and recruiting adult patients diagnosed with type 2 diabetes and residing in Douglas or Sarpy counties. RNs, who acted as health coaches, recruited eligible participants while they were still in the hospital or by telephone if they had been discharged. Patients were also invited to participate via messages sent through Nebraska Medicine's patient portal. Preliminary analyses of Medicare FFS participants in the RIISCC program indicated that 89 percent had an inpatient or outpatient stay with a primary or secondary diagnosis of type 2 diabetes at one of the two intervention hospitals within one year before enrollment.[7, 8] We have not succeeded in accounting for the absence of an inpatient or outpatient stay for the remaining 11 percent of beneficiaries. The discrepancy may be due in part to inconsistencies between Medicare claims and the hospitals' EMRs.

Because a diagnosis of type 2 diabetes and an outpatient or inpatient stay are the two enrollment criteria that can be observed in claims, we restricted our analysis to these participants and refer to the most recent stay as the *anchor encounter*. We defined the Medicare treatment group as those beneficiaries who were discharged home from an inpatient or outpatient stay with all of the following characteristics:

[7] We defined FFS beneficiaries based on two criteria: (1) enrolled in Medicare FFS (both Parts A and B), with Medicare as the primary payer on their program enrollment date, and (2) enrolled in Medicare FFS (both Parts A and B) for a period of at least 90 days during the baseline year.

[8] For treatment group beneficiaries, we defined enrollment date based on the enrollment date provided by the awardee. However, if the enrollment date was before the discharge date of the anchor encounter, we defined it as the day after the discharge date of the anchor encounter.

- The stay occurred at Nebraska Medicine or Nebraska Medicine-Bellevue hospital.
- The discharge occurred within one year before enrollment.
- The Medicare claim for the stay contained a primary or secondary diagnosis of type 2 diabetes.

We further restricted the study sample to enrolled beneficiaries who were alive at least 30 days after their enrollment date. We also limited the sample to those who enrolled on or before November 30, 2016, to allow for claims run-out for the first two quarters of the intervention period. As of November 30, 2016, we identified an anchor encounter for 351 Medicare FFS beneficiaries who participated in the RIISCC program.

The participation rate among eligible beneficiaries discharged from Nebraska Medicine or Nebraska Medicine-Bellevue hospital was less than 5 percent. The low rate was probably due in large part to the unexpected difficulty of recruiting patients after they had returned home from a hospital stay. Table III.4 displays characteristics of beneficiaries in the eligible nonparticipants sample with those of program participants, by type of stay (inpatient or outpatient). Within inpatient and outpatient strata, participants are broadly similar to eligible nonparticipants. Participants are more likely to be female and to be dually enrolled in Medicare and Medicaid. Prior-year Medicare spending for participants who entered the program after an inpatient stay was about 25 percent lower for participants than for nonparticipants. The comparison group drawn for the impact evaluation closely matches the treatment group as we demonstrate in Section 3 below. Nonetheless, the differences between the program participants and eligible nonparticipants, particularly for beneficiaries who entered the program via an inpatient stay, suggest that caution may be necessary in drawing inferences about the broader population.

Table III.4. Characteristics of enrolled and eligible nonparticipants, Nebraska Medicine

	Enrolled		Eligible nonparticipants	
Anchor encounter	Inpatient	Outpatient	Inpatient	Outpatient
Number	101	250	968	7,253
Age	65	67	67	68
Dual (%)	29	31	24	22
Female (%)	59	58	46	52
HCC score	2.35	1.87	2.83	1.83
Total spending one year before anchor encounter	30,002	18,256	38,242	19,654

HCC = hierarchical condition category.

2. Identifying a potential comparison group

The pool of potential comparison group beneficiaries consisted of Medicare FFS beneficiaries who had an inpatient or outpatient stay with a primary or secondary diagnosis of type 2 diabetes at any of four comparison hospitals in Omaha. The four comparison hospitals operate in the same service area as the intervention hospitals. We also restricted the comparison group to those who lived in the geographic area served by the two intervention hospitals, defined

by beneficiary residence Social Security Administration standard county code. We applied the following criteria to construct the Medicare comparison group for propensity score matching and outcome analysis:

- Had an inpatient or outpatient encounter with a primary or secondary diagnosis of type 2 diabetes at one of the four comparison hospitals in Omaha: CHI Creighton, Nebraska Methodist Hospital, CHI Bergan Mercy, or CHI Health Immanuel
- The encounter was between January 1, 2014, and November 30, 2016
- Discharge destination was home
- Lived in one of the counties serviced by the intervention hospitals: Douglas or Sarpy
- Alive for a period of at least 30 days after discharge

As of November 30, 2016, we identified 53,887 eligible encounters from 11,912 beneficiaries who met the eligibility criteria for the comparison group. We performed propensity score matching at the beneficiary-encounter level by using these encounters as potential anchor encounters for the comparison group beneficiaries.

3. The matching process

a. Variables for matching

We included a number of beneficiary characteristics as covariates in the propensity score model, including demographic characteristics, health status, and receipt of at least one A1c testing during the baseline year, defined as the 365 days preceding the day after discharge from the anchor encounter (Table III.5). Baseline measures of utilization and expenditure are strong predictors of future spending and therefore also included as covariates. In the preliminary analyses of the FFS participants in the RIISCC program, we observed an upward trend in average PBPM total expenditure in the last baseline quarter, indicating that the beneficiaries targeted for the intervention are high-cost individuals who used acute care services extensively in the quarter that immediately preceded enrollment. In order to take this trend into consideration, we constructed two PBPM total expenditure measures, one for the fourth baseline quarter and the other for the other three quarters in the baseline year. We used five measures for exact matching: age category, sex, anchor encounter type, dual Medicare-Medicaid eligibility, and discharge quarter of the anchor encounter.

Table III.5. Variables used for propensity score matching model, Nebraska Medicine

Variable
Exact-matching variables
Age category (< 55, 55–64, 65–69, 70–79, > 79)
Sex
Anchor encounter type (inpatient versus outpatient)
Dual Medicare-Medicaid eligibility
Discharge quarter of the anchor encounter
Regular matching variables
Race (white versus non-white)
Baseline A1c testing
Baseline HCC score[a]
Total Medicare expenditure per month in the fourth baseline quarter[b]
Total Medicare expenditure per month in the other three baseline quarters[b]
Baseline year number of acute hospital admissions[c]
Baseline year number of outpatient ED visits[d]
Comorbidities (diabetes with acute complications, diabetes with chronic complications, morbid obesity, major depression, heart failure, specified heart arrhythmias, vascular disease, chronic obstructive pulmonary disease [COPD], acute renal failure, hypertension, and hyperlipidemia)

[a]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

[b]We calculated total Medicare expenditures for the baseline year from all claims for each participant with at least one eligible day during that year.

[c]The hospitalization measure includes acute care hospital admissions and excludes all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[d]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

CMS = Centers for Medicare & Medicaid Services; ED = emergency department; ESRD = end-stage renal disease; HCC = hierarchical condition category.

For each comparison group beneficiary in a matched set, we defined the start of the follow-up period as the discharge date plus the elapsed time from discharge to enrollment for the treatment group beneficiary in the matched set. This ensures that the gap between discharge and enrollment is similar for the treatment and comparison groups.

b. Matching results

We performed optimal matching with replacement, permitting matched sets comprising no more than five matched comparisons per treated beneficiary (1T:5C). The matching process yielded a total of 1,581 potential anchor encounters from 1,374 comparison group beneficiaries, matched to the 351 treatment group cases, resulting in an average of roughly 5 comparisons per treatment group beneficiary. We further de-duplicated the comparison cases to the beneficiary level by randomly keeping only one potential anchor encounter for each comparison group beneficiary. Our final set of matched treatment and comparison group beneficiaries includes

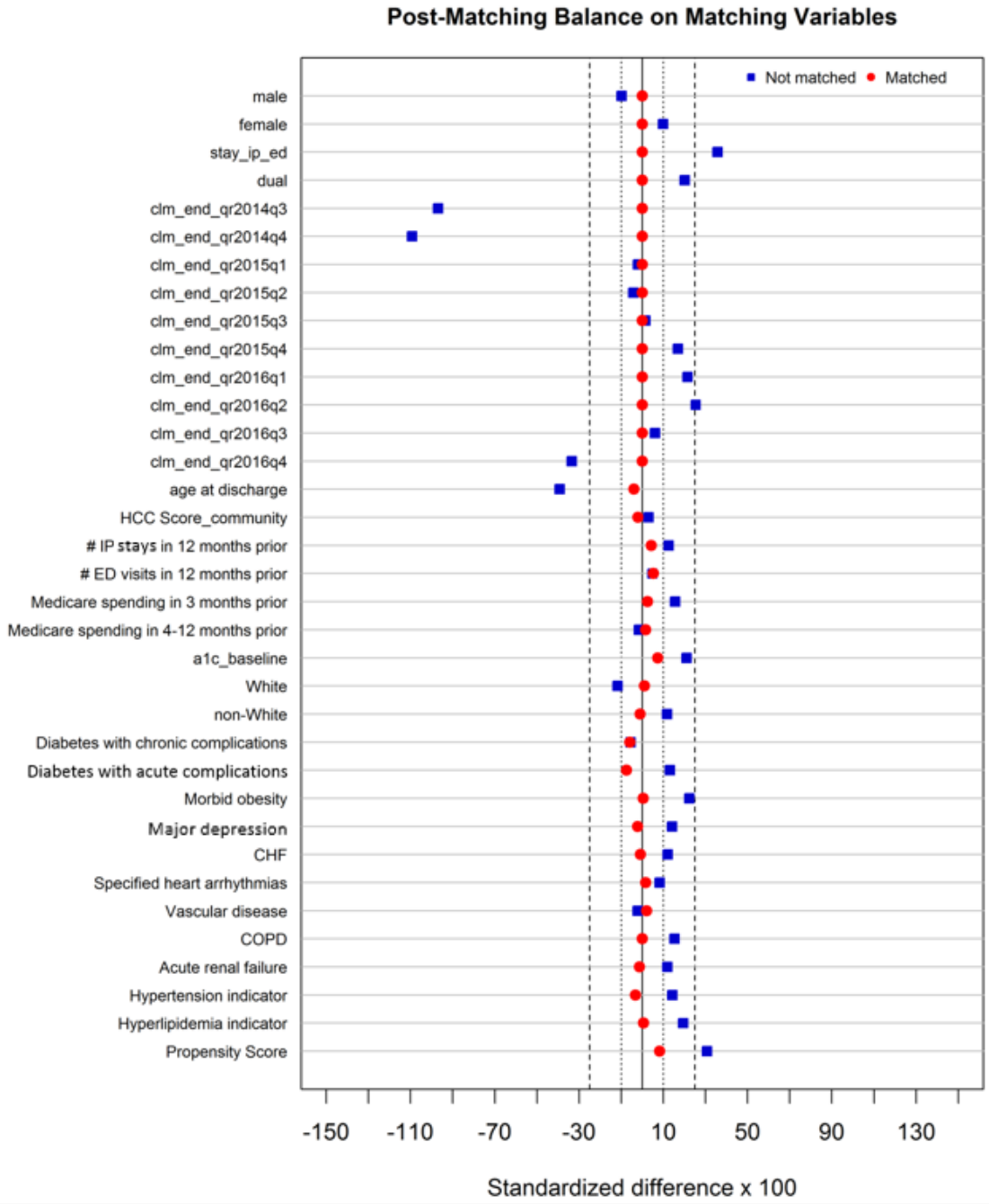
1,374 comparison group beneficiaries matched with 351 treatment group beneficiaries—an average of roughly 4 comparisons per treatment group beneficiaries, with one anchor encounter for each beneficiary. The distribution of treatment group beneficiaries by matching ratio (number of treatment group beneficiaries matched to a comparison group beneficiary) is shown in Table III.6.

Figure III.2 displays the standardized differences between the treatment and matched comparison samples for the RIISCC program.[9] Standardized differences measure the difference in weighted means between the treatment group and the matched comparison group on the standard deviation scale; in our implementation, we used the standard deviation in the treatment group to ensure that pre- and post-matching diagnostics are on the same scale and align with the evaluation's focus on impacts in the treatment group. Standardized differences between the treatment and matched comparison groups for all covariates were less than 0.10 for every matching variable.

Given the closeness of the means for treatment and comparison group beneficiaries on all variables, we believe that this matched group will yield credible estimates of program effects. The impact regression analyses will control for the remaining differences between the two groups on these characteristics.

[9] A full balance table is included in Appendix A.

Figure III.2. Comparison of balance on matching variables before versus after matching: Nebraska Medicine



Notes: We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the algorithm for new enrollees. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

A full balance table is in Appendix A of this report.

CHF = congestive heart failure; CMS = Centers for Medicare & Medicaid Services; COPD = chronic obstructive pulmonary disease; ED = emergency department; ESRD = end-stage renal disease; HCC = hierarchical condition category; IP = inpatient.

Table III.6. Distribution of Nebraska Medicine's treatment group beneficiaries, by matching ratio

Matching ratio [treatment to comparison]	Number of treatment group beneficiaries	Number of comparison group beneficiaries
1:5	123	615
1:4	123	492
1:3	70	210
1:2	22	44
1:1	13	13
Total sample size	351	1,374

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Nebraska Medicine was still developing the model at the end of the third year of the cooperative agreement. It has recently received claims data from CMS and started to conduct a study of utilization and costs and plans to use those results to engage with payers.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: (1) the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor and (2) narratives and payment model reports the awardee submitted to the implementation and monitoring contractor. We did not interview the awardee about its payment model because it did not advance from the development stage to active negotiations with one or more payers.

B. Description of the payment model

Nebraska Medicine has continued its effort in developing a bundled payment model. In this model, the payer reimburses the hospital a set amount per patient for one episode of services (90-day RIISCC intervention) provided by the whole RPM team—including PCPs, nurses, MAs, dietitians, ophthalmologists, and other staff—rather than paying for each individual service. The awardee describes the approach as consistent with Model 3 in the Bundled Payments for Care Improvement initiative, in which Medicare continues to make FFS payments and reconciles the payment later based on a bundled payment amount determined by CMS. Based on the difference between the aggregate expenditures and the target price, Medicare then makes a payment or recoupment.

To determine a reasonable charge rate for telemedicine services provided to patients with diabetes, the awardee conducted a preliminary analysis to estimate the cost of providing the telemedicine services to a patient during the 90-day RIISCC intervention, using Nebraska Medicine's internal claims data. The awardee estimated a per patient-day cost of $17 to $22 based on its direct and overhead costs in relation to the total number of patients the program serves. It plans to conduct payer-specific analyses by using claims data recently received from Medicare and Nebraska Blue Cross and Blue Shield.

C. Status of the payment model

Nebraska Medicine was still developing the model at the end of the third year of the cooperative agreement. It had recently received claims data from CMS and started to conduct a study of utilization and costs and plans to use those results to engage with payers. Nebraska Medicine also identified a comparison sample of patients to assess the effectiveness of the program in reducing health care utilization and costs relative to similar patients who receive conventional care. This information is important for engaging and motivating payers to participate in the new payment model. Its initial results indicated that when compared to the nonparticipants, program participants had substantially lower number of days of inpatient care and related costs after the 90-day RPM intervention. The awardee presented these findings internally and has been using feedback to prepare for meetings with Nebraska Medicine leaders and potential payers.

The awardee leaders have initiated a conversation with Nebraska Blue Cross and Blue Shield. In addition, they are working with the payer relations department at Nebraska Medicine in the hopes that it would be more successful in starting a dialogue with the state Medicaid agency and private payers. However, awardee leaders do not expect implementation of the new payment model by the end of the HCIA R2 cooperative agreement.

D. Factors associated with the development of the payment model

The awardee's payment model has been hindered by the lack of claims data from payers other than Blue Cross Blue Shield. The awardee submitted data use agreement forms and supporting documents to CMS to access Medicare data but did not receive the data until May 2017, near the end of the third year of the cooperative agreement. It also reached out to United Health Care and the Nebraska Medicaid office regarding accessing claims data for patients who participated in the RPM intervention, but was still waiting to receive data from these payers.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Nebraska Medicine focused on engaging organizational leaders as its sustainability strategy in the third program year. The awardee began outlining the program's benefits relative to its cost and set plans to meet with organizational leaders and payers to discuss funding and sustaining the program.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Nebraska Medicine was developing a bundled payment model with shared savings, based on service episode. Nebraska Medicine also reported plans to discuss reimbursements for telehealth services from some payers; the reimbursements will provide a small revenue source to sustainably fund the program. Finally, the awardee focused on (1) engaging stakeholders by analyzing program outcome data to show that the program improved outcomes, and (2) improving efficiency of program operations.

C. Implementing the SSR plan: progress and changes

Sustainability. By the end of the third program year, Nebraska Medicine focused on engaging organizational leaders to sustain the program. First, the awardee documented the program's costs, as well as the financial impact on health care spending attributable to the program. Awardee leaders expressed belief that savings from the program relative to its cost will help engage organizational leaders. Awardee leaders have met, and will continue to meet, with organizational leaders to discuss options to sustain the program and potentially expand the target population to other chronic care populations. The awardee received guidance on submitting a claim for state Medicaid's reimbursement of these services. Awardee leaders also reported plans to engage commercial payers to see whether they would provide payments for any program services. Nebraska Medicine recently started to provide telehealth services to its patients based on clinician referrals (rather than recent hospitalizations). These services are being sustained without program funding.

Scalability. Nebraska Medicine did not report plans to scale its program.

Replicability. Nebraska Medicine did not report plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

Nebraska Medicine expressed hope that the program's positive outcomes related to cost savings will engage organizational leaders and payers. Low numbers of participants insured by payers other than Medicare and Medicaid prevented the awardee from showing costs and cost savings by payer, delaying the awardee in finalizing details about sustaining its program. In the third program year, Nebraska Medicine hired a consultant team to help gain support from organizational leaders and payers to sustain the program by analyzing program data, providing guidance on reimbursement opportunities, and developing messaging.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. However, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or the SSR plans. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

With matching now complete, we plan to begin developing the first round of impact estimates. These estimates will focus on standard outcomes such as hospitalization, readmission, ED use, and Medicare spending over two successive six-month periods after program enrollment. For each outcome, we will report the impact estimate, the 90-percent confidence interval for the estimate, and the *p*-value associated with the null hypothesis that the impact is zero. We will report these estimates to CMS as they become available.

This page has been left blank for double-sided copying.

APPENDIX A

POST-MATCHING DIAGNOSTICS

This page has been left blank for double-sided copying.

Table A.1. Post-matching diagnostics

Measure	Potential comparisons (N = 44,803)	Matched comparisons (N = 1,394)	Treated (N = 351)	Adjusted difference (Matched)	Percentage difference (Matched)	Standardized difference (Matched)	*p*-value (Matched)	Equivalence *p*-value (Matched)
Proportion male	0.462	0.413	0.413	0.000	0.000	0.000	1.000	0.000
	(0.002)	(0.013)	(0.026)	(0.036)				
Proportion female	0.538	0.587	0.587	0.000	0.000	0.000	1.000	0.000
	(0.002)	(0.013)	(0.026)	(0.036)				
Proportion with inpatient anchor stay	0.200	0.373	0.373	0.000	0.000	0.000	1.000	0.000
	(0.002)	(0.013)	(0.026)	(0.036)				
Proportion dual eligible	0.212	0.305	0.305	0.000	0.000	0.000	1.000	0.000
	(0.002)	(0.012)	(0.025)	(0.034)				
Proportion with anchor stay in 2014 Q3	0.115	0.011	0.011	0.000	0.000	0.000	1.000	0.001
	(0.002)	(0.003)	(0.006)	(0.008)				
Proportion with anchor stay in 2014 Q4	0.109	0.009	0.009	0.000	0.000	0.000	1.000	0.001
	(0.001)	(0.002)	(0.005)	(0.007)				
Proportion with anchor stay in 2015 Q1	0.106	0.100	0.100	0.000	0.000	0.000	1.000	0.000
	(0.001)	(0.008)	(0.016)	(0.022)				
Proportion with anchor stay in 2015 Q2	0.112	0.100	0.100	0.000	0.000	0.000	1.000	0.000
	(0.001)	(0.008)	(0.016)	(0.022)				
Proportion with anchor stay in 2015 Q3	0.101	0.105	0.105	0.000	0.000	0.000	1.000	0.000
	(0.001)	(0.008)	(0.016)	(0.023)				
Proportion with anchor stay in 2015 Q4	0.093	0.154	0.154	0.000	0.000	0.000	1.000	0.001
	(0.001)	(0.010)	(0.019)	(0.028)				
Proportion with anchor stay in 2016 Q1	0.097	0.179	0.179	0.000	0.000	0.000	1.000	0.000
	(0.001)	(0.010)	(0.021)	(0.029)				
Proportion with anchor stay in 2016 Q2	0.100	0.202	0.202	0.000	0.000	0.000	1.000	0.000
	(0.001)	(0.011)	(0.021)	(0.029)				
Proportion with anchor stay in 2016 Q3	0.100	0.120	0.120	0.000	0.000	0.000	1.000	0.000
	(0.001)	(0.009)	(0.017)	(0.024)				
Proportion with anchor stay in 2016 Q4	0.067	0.020	0.020	0.000	0.000	0.000	1.000	0.001
	(0.001)	(0.004)	(0.007)	(0.011)				
Age at discharge	70.364	66.774	66.368	-0.407	-0.600	-0.040	0.611	0.004
	(0.055)	(0.275)	(0.543)	(0.800)				

Table A.1 *(continued)*

Measure	Potential comparisons (N = 44,803)	Matched comparisons (N = 1,394)	Treated (N = 351)	Adjusted difference (Matched)	Percentage difference (Matched)	Standardized difference (Matched)	*p*-value (Matched)	Equivalence *p*-value (Matched)
HCC Score_community	1.973	2.037	2.011	-0.026	-1.300	-0.020	0.804	0.002
	(0.008)	(0.038)	(0.069)	(0.104)				
# IP stays in 12 months prior	0.765	0.902	0.973	0.071	7.300	0.043	0.563	0.003
	(0.011)	(0.040)	(0.089)	(0.123)				
# ED visits in 12 months prior	1.233	1.210	1.420	0.210	14.800	0.054	0.383	0.001
	(0.019)	(0.058)	(0.209)	(0.240)				
Medicare spending in 3 months prior	6054.991	8087.009	8474.805	387.796	4.600	0.025	0.724	0.001
	(58.393)	(334.949)	(831.109)	(1098.006)				
Medicare spending in 4–12 months prior	12449.217	11437.162	11949.469	512.307	4.300	0.016	0.807	0.000
	(113.756)	(565.159)	(1701.695)	(2095.335)				
Proportion with a1c test at baseline	0.791	0.838	0.863	0.025	2.900	0.074	0.357	0.014
	(0.002)	(0.010)	(0.018)	(0.027)				
Proportion white	0.800	0.745	0.749	0.005	0.600	0.011	0.882	0.001
	(0.002)	(0.011)	(0.023)	(0.032)				
Proportion non-white	0.200	0.255	0.251	-0.005	-1.900	-0.011	0.882	0.001
	(0.002)	(0.011)	(0.023)	(0.032)				
Diabetes with chronic complications (proportion)	0.021	0.021	0.014	-0.007	-49.300	-0.059	0.477	0.011
	(0.001)	(0.004)	(0.006)	(0.010)				
Diabetes without complication (proportion)	0.546	0.646	0.610	-0.037	-6.000	-0.075	0.327	0.011
	(0.002)	(0.013)	(0.026)	(0.037)				
Morbid obesity (proportion)	0.120	0.209	0.211	0.002	0.900	0.005	0.951	0.001
	(0.002)	(0.010)	(0.022)	(0.031)				
Major depression (proportion)	0.108	0.168	0.160	-0.008	-5.100	-0.022	0.766	0.001
	(0.001)	(0.010)	(0.020)	(0.027)				
CHF (proportion)	0.225	0.283	0.279	-0.004	-1.400	-0.009	0.911	0.001
	(0.002)	(0.012)	(0.024)	(0.034)				
Specified heart arrhythmias (proportion)	0.221	0.249	0.256	0.007	2.900	0.017	0.820	0.001
	(0.002)	(0.012)	(0.023)	(0.032)				
Vascular disease (proportion)	0.200	0.183	0.191	0.008	4.300	0.021	0.792	0.002
	(0.002)	(0.010)	(0.021)	(0.031)				

Table A.1 *(continued)*

Measure	Potential comparisons (N = 44,803)	Matched comparisons (N = 1,394)	Treated (N = 351)	Adjusted difference (Matched)	Percentage difference (Matched)	Standardized difference (Matched)	*p*-value (Matched)	Equivalence *p*-value (Matched)
COPD (proportion)	0.205	0.274	0.274	0.000	0.000	0.000	1.000	0.000
	(0.002)	(0.012)	(0.024)	(0.032)				
Acute renal failure (proportion)	0.124	0.173	0.168	-0.005	-2.800	-0.013	0.873	0.001
	(0.002)	(0.010)	(0.020)	(0.029)				
Hypertension indicator (proportion)	0.896	0.940	0.932	-0.008	-0.900	-0.033	0.650	0.001
	(0.001)	(0.007)	(0.013)	(0.018)				
Hyperlipidemia indicator (proportion)	0.739	0.812	0.815	0.002	0.300	0.006	0.935	0.001
	(0.002)	(0.011)	(0.021)	(0.030)				
Propensity score	0.008	0.011	0.012	0.001	9.500	0.082	0.166	0.002
	(0.000)	(0.000)	(0.001)	(0.001)				

	Chi-squared statistic	Degrees of freedom	*p*-value
Omnibus test for balance on matching variables	19.279	19.000	0.439

Notes: Standard errors in parentheses. Standardized difference calculated as the difference in means divided by the treatment group standard deviation. *P*-values come from a weighted two-sample t-test. The equivalence test comprises two one-sided tests of whether the difference between the treatment and comparison group means exceeds 0.25 standard deviations (in either direction) of that variable. If we can reject that hypothesis, we are confident that the true difference between the two populations being compared is less than 0.25 standard deviations.

HCC = hierarchical condition code; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE,WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Northwell Health

April 18, 2018

Craig Schneider (Mathematica Policy Research)
Lisa Lines (RTI International)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Northwell Health, formerly North Shore–Long Island Jewish Health System, used funding from HCIA R2 to create the Healthy Transitions program, which included hiring nurse care

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Northwell Health was planning to enroll and serve patients until the end of the no-cost extension because it intended to continue the program beyond the extension with external funding.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

managers, and expand the number of implementing sites beyond the Northwell system to community nephrology practices in four counties.

Too often, patients with advanced kidney disease (Stages 4 and 5)[4] "crash" and arrive at the ED in urgent need of dialysis. The standard emergency treatment is to insert a dialysis catheter in order to immediately start hemodialysis. However, dialysis catheters have a high risk of infection and other medical problems. The goal of the Healthy Transitions model was to intervene early enough in the kidney disease progression to enable patients to make appropriate choices about their treatment options, which include kidney transplantation; different kinds of dialysis (hemodialysis versus peritoneal dialysis); possible surgery to create an arteriovenous fistula (AV fistula) for hemodialysis, which has many advantages over dialysis catheters; or to forego transplantation or dialysis and seek comfort care only (also called "conservative care"). The Healthy Transitions model included home visits by nurse care managers to determine whether the patient had sufficient family support and an adequate home environment. In addition, they (1) educated patients about diet and exercise to slow the progress of the disease and delay the start of dialysis for as long as possible and (2) promoted transplantation, which is the optimal outcome for those who do not choose conservative care.

The Healthy Transitions processes and systems were mostly in place prior to the HCIA R2 award. During the course of the cooperative agreement, the awardee upgraded its computer tracking systems and improved its program model based on lessons learned during implementation. By program Year 3, Northwell expected to have 13 sites participating in the Healthy Transitions program, including the original Northwell Great Neck site.

Table I.1 presents additional details about the Healthy Transitions program and model.

[4] Chronic kidney disease is categorized into five stages, with Stage 5 being the most severe. See https://www.kidney.org/atoz/content/gfr.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Healthy Transitions is a patient-centered program that aims to integrate and coordinate all aspects of care for persons with late-stage chronic kidney disease (CKD) by (1) focusing on patient education and care management to delay the onset of end-stage renal disease (ESRD) and (2) helping patients make informed choices about ESRD treatment that reflect their personal preferences.
Major innovation	Intervening early enough in the disease process to enable patients to make appropriate decisions about their CKD treatment options.
Program components	• Care management • Shared decision making
Target population	Eligible participants are individuals with late-stage CKD (Stages 4 and 5) who meet the following requirements: (1) are at least 18 years old, (2) live in one of four specified NY counties, (3) have an estimated glomerular filtration rate of less than 30 ml/min, and (4) have no clinically apparent cognitive impairment.
Theory of change/ theory of action	Northwell focuses on changing participant and provider behavior by shifting the nephrologist-based care model to a greater reliance on nurse care managers, who can develop more personal relationships with the participants and guide them through the complex care system. The awardee believes that this improved model of disease management will lead to improved outcomes and better preparation for ESRD, which will ultimately lower costs.
Payment model	Value-based payments, bundled or episode payment, and capitated payment for care management and care coordination services
Award amount	$2,453,742
Effective launch date[a]	November 17, 2014
Program setting	Patients' homes, practice offices
Market area	Urban, suburban
Market location	Manhattan, Nassau, Queens, and Suffolk counties in NY
Target outcomes	• Patients better prepared for ESRD care • Increased patient selection of modality • Increased AV fistula rate • Increased quality of life score • Savings to Medicare of $1.9 million (based on the original estimate of 593 participants)

[a]After the initial planning period, the awardee's program became operational as of this date.

AV = arteriovenous.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we conclude that Northwell was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on six factors. First, the awardee enrolled 629 participants—126 percent of its target enrollment goal—by the end of the initial cooperative agreement, which means that Northwell exceeded the level for achieving enrollment effectiveness. Second, the awardee was largely able to effectively deliver its intervention services by using nurse care managers to engage physicians and patients, although geographic

expansion was more limited than initially planned. Third, the program was partly successful in meeting its staffing and training goals. Although training was implemented and the program was fully staffed overall, key vacancies prevented expansion to all planned sites and resulted in suspension of one of the modality options. Fourth, Healthy Transitions was successful in recruiting and engaging providers, based on feedback provided by interview respondents. Fifth, the awardee was successful in engaging participants, as indicated by survey and interview responses. Finally, participating clinicians and other implementation staff perceived the program as having a positive effect on care delivery, according to our interview and survey data and the 12th quarter survey conducted by the implementation and monitoring contractor.

Impact evaluation. Due to too few treatment beneficiaries, we do not anticipate being able to conduct a rigorous impact analysis for Northwell's Healthy Transitions program. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. Northwell's proposed payment model combines a condition-specific, population-based payment for people with chronic kidney disease (CKD)—before end-stage renal disease (ESRD)—with a small set of quality-based incentive or penalty payments for nephrologists.

Sustainability plans. As of the third program year, Northwell was continuing to look for sources of external funding to sustain its program. Northwell did not report activities to scale its program during the third program year. The awardee held conversations with a few Medicare ESRD Seamless Care Organizations (ESCOs) about replicating Healthy Transitions, but none of these potential initiatives were finalized by the end of the cooperative agreement.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey, which was fielded from July 2016 to October 2016, achieved a response rate of 100 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter were determined by the following measures: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals. We found that, although the program reached its revised enrollment goal and seems to have improved quality of care and efficiency, staffing issues prevented expansion into Manhattan and Brooklyn and limited the availability of the conservative care option during the third program year.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

The program, which began in November 2014, focused on managing issues associated with CKD, such as metabolic complications, comorbidities, the burdens of hospitalization, and preparation for treatment of ESRD. To implement the program, Northwell partnered with six nephrology practices in four counties in the New York City area (Manhattan, Nassau, Queens, and Suffolk counties). Many of the practices operated at multiple locations. Healthy Transitions was expected to be available at 13 sites. However, the Manhattan sites were never implemented. By the end of the third program year, Healthy Transitions was operational at 9 sites.

The target population was patients with advanced CKD—that is, patients in Stages 4 and 5. The program excluded individuals who reached ESRD or were on dialysis. To be eligible for the program, participants had to be covered by a nonmilitary payer (Medicare, Medicaid, or commercial insurance) and meet the following criteria: (1) were at least 18 years old, (2) lived in one of the four specified New York counties, (3) had an estimated glomerular filtration rate (eGFR) of less than 30 ml/min,[5] and (4) had no clinically apparent cognitive impairment. There was no cost for participants.

Program leaders projected that Healthy Transitions would serve 500 participants over the three years of the cooperative agreement. Their goal was to have 400 patients enrolled at any one time, but they said it would take about 570 patients overall to reach this level because so many enrollees lose program eligibility due to progression to ESRD, election of hospice services, or death. The awardee's goals for the program were to (1) improve patient education and shared decision making to better prepare patients for ESRD care; (2) increase the percentage of individuals with advanced kidney disease who actively choose and prepare for home dialysis modalities and preemptive kidney transplants in order to improve their quality of life; (3) increase the rate of AV fistulas; (4) increase the quality of life score; and (5) generate savings to Medicare of $1.9 million (based on an original estimate of 593 participants), largely through reductions in hospitalizations, ED utilization, and costly and harmful delays of dialysis treatment.

During the cooperative agreement, two important changes were made to recruitment and enrollment processes: (1) nurse care managers were embedded in the nephrologists' offices, where they offered printed materials to promote enrollment in the program, and (2) nurse care managers sent reminders to nephrologists about upcoming patients who may have been eligible for the program.

b. Evidence of enrollment effectiveness

Overall, Northwell reported that it enrolled 629 participants from November 2014 (when it launched its program) through August 2017—about 126 percent of its 500 three-year projected participants[6] (Figure II.1). Enrollment increased from 34 percent of the goal (170 participants) at the end of Year 1 to 79 percent (395 participants) by the end of Year 2 to 126 percent (629 participants) by the end of Year 3. Of the 629 participants, 372 participants remained in the program at the end of the cooperative agreement, while 257 disenrolled due to progression to ESRD (191), death (30), election of hospice care (12), and miscellaneous reasons such as moving out of the geographic area (26).

Northwell sent reminders to the participating nephrologists that patients whose eGFR dropped below 30 ml/min may be eligible for the program. During the second program year, Northwell began embedding nurse care managers in the nephrologists' offices. Program leaders

[5] The eGFR is a blood test that indicates how well the kidneys are filtering. An eGFR of 60 ml/min or higher is in the normal range, below 60 ml/min signifies kidney disease, and 15 ml/min or lower means kidney failure. See https://www.niddk.nih.gov/health-information/health-communication-programs/nkdep/learn/causes-kidney-disease/testing/understand-gfr/Pages/understand-gfr.aspx.

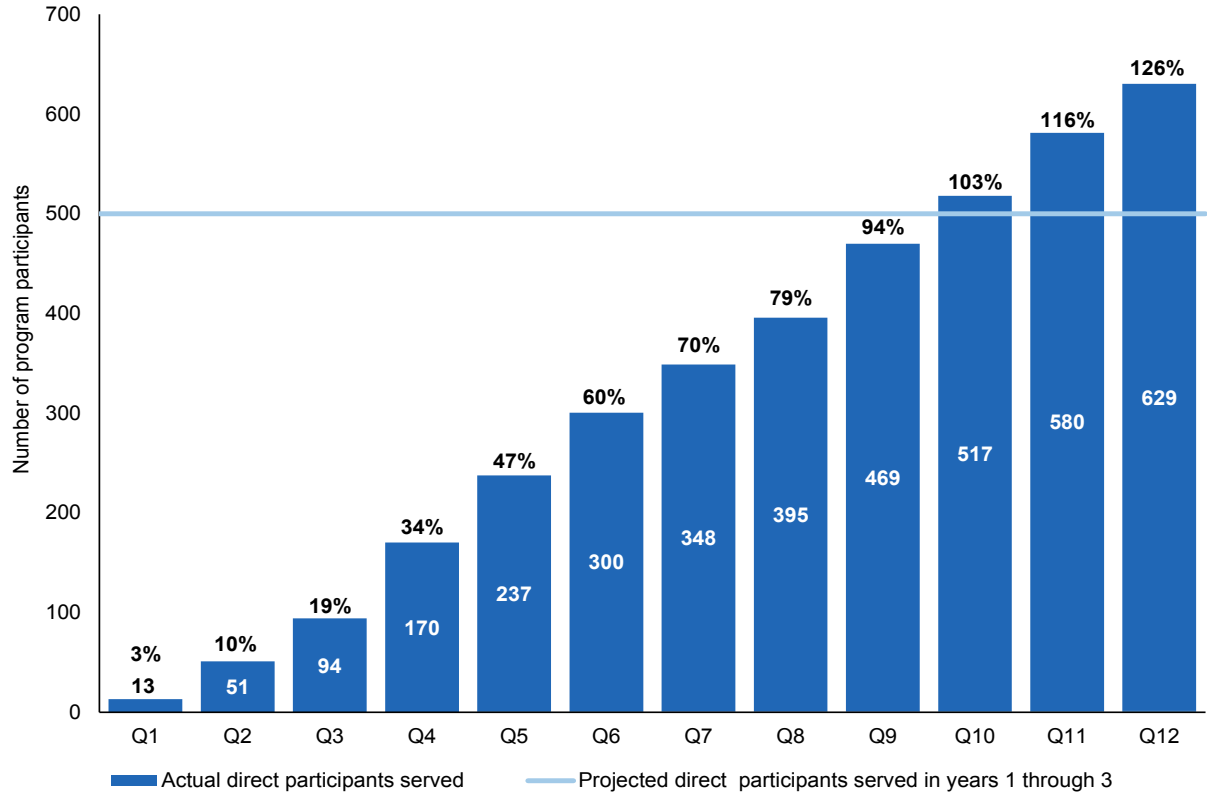
[6] The awardee lowered its enrollment projections from 593 participants to 500 participants during Year 2.

believed that this approach increased not only program enrollment but also patient trust in the nurse care managers. Upon approval from the nephrologist, the nurse care manager approached potential participants about enrolling in the program during their scheduled visit to the nephrologist.

In addition, the support the nurses provided to the nephrologists increased their willingness to refer patients to the program. One nephrologist called the nurse care managers' presence "invaluable." In addition, one nurse care manager believed that having the nurses present in the clinics increased trust between the patients and the nurse care managers. Having the nurse care manager on-site at the time of the nephrologist's consultation with the patient also enabled questions about the program to be answered immediately and assisted with same-day enrollment.

> "I believe that [being embedded in the nephrologists' offices] has increased our enrollment, definitely. Just being there and being available to the nephrologist."
>
> —Nurse care manager

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding.

c. Barriers and facilitators associated with enrollment effectiveness

Northwell's progress in meeting its three-year enrollment goal was influenced by several factors. One facilitator was that the program was free. One nephrologist pointed out that the first question patients asked was whether there was an extra cost for participating in the program. Physician engagement was another facilitator. Another nephrologist noted that there was a "very active approach" to engage physicians in Healthy Transitions, which included the program providing reminders to them about potentially eligible patients. The participating nephrologists recognized the value of the program both to themselves and to their patients, and therefore wanted to enroll patients in the program. An additional facilitator was the dedication and commitment of the nurse care managers, who spent an extensive amount of time communicating the benefits of the program to potential enrollees.

However, there continued to be barriers to enrollment through the third program year. One barrier was the enrollment of high-acuity patients. Patients enrolling too late in the disease process did not have sufficient time for the intervention—that is, education, shared decision making, surgery to implant an AV fistula, recovery from the surgery, and the care coordination components—to be effective before their kidneys completely failed and they faced the urgent need for dialysis. Despite this, practices continued to refer patients with very low eGFR levels (that is, patients with advanced kidney disease on the verge of needing dialysis) and the program accepted them because the staff felt that the nurse care managers' services would benefit the patients, even though these patients would negatively affect the program's results. Another barrier was staff turnover, in particular the departure of the nurse who was to be assigned to the Manhattan territory. Without a nurse to visit patients' homes and be present in the practices, enrollment ceased in Manhattan and the planned expansion to Brooklyn was put on hold.

2. Delivery of program services

a. Description of and changes to the service delivery model

The theory of action for the model was that a trained nurse care manager was assigned to serve patients who lived within specified zip codes, and then embedded in a participating nephrologist's practice to be the "face" of the program for both the nephrologist and the patients. The nurse care managers identified potential program participants by using patients' charts (that is, by looking for eGFR levels < 30 ml/min). They then conferred with the nephrologists about the patients' eligibility and determined whether the patients were appropriate for the program. Once the patient was enrolled, the nurse care manager conducted a home visit within two weeks to assess the patient's home environment and social support needs. The home visit was also an opportunity for the nurse care managers to teach the patients about their treatment options and how to manage their disease through diet, exercise, and medication.

The desired outcomes of this process were to (1) reduce ED visits and hospitalizations by monitoring daily fluctuations in patient weight, (2) enable patients to select the optimal renal replacement therapy (RRT) through shared decision making, (3) increase the proportion of participants who had actively selected RRT (the modality selection rate), (4) increase the percentage of patients choosing and preparing for home dialysis and transplant, (5) reduce the rate of starting hemodialysis with a dialysis catheter, and (6) reduce the total costs of care in late-stage kidney disease.

Program participants used a simple telephone system known as FONELink, and participants were assigned a unique identifier that they used to report their weight on a daily basis. Nurse care managers followed up when a participant's weight suddenly increased (the participants were also given scales by Northwell to track their weight). In addition, the nurses tried to follow up with participants by phone once a month. These calls included ongoing medication reconciliation. For high-risk patients, the phone calls were conducted weekly. Follow-up home visits were scheduled on an annual basis, unless the nurse care manager determined that another home visit should be conducted sooner.

FONELink data were collected and reported in a daily tracking report, which was sent to each nurse care manager. The daily report included the status of the participant's treatment option decision, weight trend for the past 10 days, upcoming and previous scheduled visits, most recent eGFR with date of service, risk stratification, and any notes or comments on the patient.

Following the home visit, the nurse care manager monitored and tracked participants' disease progression, continued to educate participants about the disease, and linked them to needed social supports or to the social worker. When a participant's eGFR level dropped below 20 ml/min, the care manager began to work with the participant on choosing an RRT modality and preparing for RRT. When the participant's eGFR levels dropped below 15 ml/min, the participant was considered to be in ESRD and began dialysis or conservative care management, at which point the participant was disenrolled from the program.

During the cooperative agreement, several changes were made to the service delivery model. In addition to embedding nurse care managers in the nephrologists' offices and sending reminders about potentially eligible patients, the most notable change to the model after implementation was the addition of the conservative care model as an RRT modality option. However, this aspect of the program was suspended at the time of our Year 3 virtual site visit due to the departure of the nephrologist with palliative care expertise, who had been leading this initiative. Another change during the third program year was that Healthy Transitions staff began meeting quarterly with the Northwell Health transplant center to encourage a team-based approach to caring for these patients. They then tracked the patients through the transplant process. Finally, the program switched to a new database system during the third program year, which enhanced speed and functionality.

b. Evidence of service delivery effectiveness

Healthy Transitions was successful on three of the four dimensions of the service delivery model—delivery of intervention services, recruitment and engagement of providers, and engagement of program participants—and partly successful on the fourth dimension of staffing and training. Although the program ultimately reached its enrollment goal and seems to have improved quality of care and efficiency, staffing issues prevented the program's expansion and limited the availability of the conservative care option during the third program year.

The awardee reported to the implementation and monitoring contractor towards the beginning of the 12th quarter that it had achieved most of its goals to improve health outcomes and quality of care and had achieved some of its goals to save costs and develop a payment model.

Delivery of intervention services. The evidence from interviews, survey data, and the awardee's reports to the implementation and monitoring contractor indicates that the awardee delivered the intervention services as intended. However, although the model described in the theory of action was delivered effectively overall, there were some challenges caused by staffing issues that prevented complete success: inability to expand to Manhattan and Brooklyn, suspension of the conservative care option, and not implementing protocols to avoid enrolling patients in later stages of the disease.

Staffing and training. Staffing and training were partly successful. The final report prepared by the implementation and monitoring contractor stated that the program hired its goal of 7 full-time equivalent staff (which was actually 11 total full-time and part-time staff members) and trained all staff members. The most significant staffing challenge was that the program lost the nurse care manager who was to be assigned to the Manhattan practices. This setback ultimately prevented program implementation at the two Manhattan practices. Although this was unfortunate, program leaders stated that these practices had a low patient census, so this development did not have an overall negative effect on enrollment. As mentioned, the palliative care nephrologist also left the program, which resulted in the suspension of the conservative care option. However, in the staff survey, nearly all respondents felt that staff turnover was either not a problem or only a minor problem.

During interviews conducted in the last half of the third program year, program leaders reported that the average caseloads of the nurse care manager ranged from 85 to 107 participants, close to what they felt was the optimal caseload of 100 participants. However, nurse care managers felt that the optimal caseload was much less (about 75 participants). They noted that the quality of their services would start to deteriorate above that level. Furthermore, they considered caseloads of 100 participants to be unmanageable. The patient satisfaction data reported by the awardee to the implementation and monitoring contractor implied that the caseload level did not affect participants' perceptions of the quality of services they received.

Nearly all survey respondents reported that they had attended a formal training session. Opinions of the value of the training were mixed, with about half of respondents feeling strongly that the training had been valuable and the rest feeling that the training had been either somewhat valuable or not valuable. Responses to the question about whether training helped improve job performance were similarly mixed. Additional training was not seen as worthwhile: about half strongly disagreed that more training would be helpful. Survey respondents reported that informal training was fairly common, including self-study; asking colleagues for help; receiving individual supervision; receiving mentoring, group supervision, or technical assistance; shadowing other staff; and participating in huddles.

Recruitment and engagement of providers. There was evidence that the multiple strategies pursued by the program to engage providers were successful. For example, Healthy Transitions provided nephrologists report cards about how their patients were doing. One nephrologist said that these reports were helpful in encouraging him to engage patients earlier in the disease process. In addition, leaders at Healthy Transitions formed a medical advisory board that met quarterly. These meetings helped program leaders and nephrologists address difficult situations with patients, including how to help them navigate the health system or assist them

> "I'm able to see more patients and manage more effectively the CKD Stage 4 patients. Because of the program, I'm able to really prepare them in a very timely fashion, on a very programmed and protocolized fashion. That made it very, very, very effective."
>
> —Nephrologist

with the RRT process. Furthermore, embedding the nurse care managers helped promote enrollment and enhanced nephrologist satisfaction.

In addition to working with the nephrologists, the nurse care managers communicated with vascular surgeons who performed the AV fistula surgeries and with the transplant teams. One nephrologist commented that communications improved "in a very comprehensive way."

Engagement of program participants. Engaging participants, beginning with the home visit, was a critical component of the Healthy Transitions model. In fact, the nurse care managers believed that the home visit was what set Healthy Transitions apart from other disease management interventions. There was some anecdotal evidence that engagement of program participants was successful. The nurse care managers gave the example of a nurse who runs a telephone-only service, but referred a patient to Healthy Transitions because she believed the home visit was more effective than her own telephone-only program.

> "I just cannot thank … the entire Healthy Transitions team enough. You have helped me navigate through an enormously complex medical situation [and] I know I will always get an expedient, intelligent response. You have brought me great peace of mind and I am extremely grateful!"
>
> —Participant

Program leaders reported that patient satisfaction exceeded 95 percent on their semiannual patient engagement surveys. The final quarterly report they submitted included several very positive patient comments.

c. Barriers and facilitators associated with service delivery effectiveness

Interview respondents identified several barriers in addition to the barriers that were discussed in the staff survey. One barrier was that Healthy Transitions continued to receive high-acuity (extremely sick) patients. The nurse care managers said that some nephrologists referred patients to the program who were on the brink of dialysis. "This makes the program go from prevention to crisis [management]," one nurse care manager said.

Another barrier was patient health literacy and lack of health education. Potential participants did not always understand the nature of their disease, sometimes had trouble communicating their preferences, or did not speak English as their first language. One aspect of not understanding the disease process was that when the nurses told the patients to begin preparing for dialysis the patients replied that they did not understand why they needed to take any steps because they felt fine. These considerations made it difficult for patients to hear the message. Clinicians reported that it could take three visits and multiple attempts to communicate the benefits of the program before the patient was ready to consider enrollment.

Socioeconomic status and the patient's home environment were also barriers to success. Some patients hesitated to enroll because they were embarrassed by what they perceived to be their humble or modest home conditions. Yet, the home visit was a program requirement. In part due to this factor, a significant percentage of potential participants (program leaders estimated 40

to 50 percent) initially declined enrollment. However, once enrolled the opt-out rate was relatively low.

In order for participants' homes to be suitable for home dialysis, they had to have enough room for the dialysis equipment, be accessible for delivery of supplies, and meet certain basic criteria for cleanliness and mobility within the home. In an interview, program leaders said that patients having an appropriate home environment was one of two barriers to increasing home dialysis rates. The other barrier was that one of the leading peritoneal dialysis fluids manufacturers ceased manufacturing, which caused the peritoneal dialysis program to grind to a halt.

Nearly all survey respondents reported that program technology and health information technology were major barriers. Other major issues cited included difficulty working across collaborating organizations and document requirements. Some of the barriers considered to be "minor" by at least half of respondents included staff turnover or unfilled positions, insufficient staff time or workload, resistance to the program by clinicians, and ineffective communication.

The facilitators for service delivery effectiveness were the new database, the collaborative role of the social worker, an improved relationship with the transplant center, and program leadership. Although there had been numerous complaints about the database and the data collection process and tool during the first two program years, interview respondents stated that, once implemented in the third program year, the new database had freed up the nurses' time and allowed them to be more efficient in enrolling patients.

Staff commented during interviews that the collaboration between the social worker and the nurse care managers was a facilitator. For example, the interaction between the social worker and the patient was sometimes used by the nurses to get access to the patients and enroll them in the program.

Another change from a former barrier to a recent facilitator was the partnership with the transplant center. Previously, the amount of time it took from issuing a referral to receiving an appointment at the transplant center was considered unacceptable. Changes in staffing at the transplant center that occurred independently of the Healthy Transitions program resulted in an improved relationship between Healthy Transitions and the transplant center. "I think our connection to the transplant program is probably better than it ever has been," a program leader said. "This is important because we do a lot of work here and it's probably the most important part of the program—getting preemptive transplantation done."

> "We basically don't let them leave [the program]. Once we made that home visit, we continue. Sometimes after a month or two they just totally ignore us, but we just keep trying. We say, 'Remember I told you when we first met that I'd be calling you once a month?' I feel like it works to remind them that I'm not making this up or pestering them, that this is my role."
>
> —Nurse care manager

The dedication of the nurses was an important factor in achieving service delivery effectiveness. When conducting home visits, one nurse care manager reported that the nurses were persistent in trying to convince the patients to participate.

Interview respondents said that another facilitator was the leadership, vision, and reputation in the CKD community of the program director. One nurse care manager said: "We've seen an incredible growth over five years [including the time prior to the cooperative agreement] of people completely changing their practice and believing in what we're doing. Initially, they hated it and hated us at first, [but] our director made sure that they really buy into all of what we're saying…. I can look back and say we've really changed the practice of [CKD care]."

C. Assessment of perceived program effects on the delivery of care and outcomes

There was consensus among program leaders, non-clinician staff, and the nephrologists whom we interviewed that the program largely achieved its objectives to improve quality of care, reduce the cost of care, and to make care for CKD patients more patient-centered.

Program leaders and staff reported that they believed that the intended outcomes (increasing the percentage of patients choosing and preparing for home dialysis and reducing the overall cost of care for late-stage kidney disease) are achievable. Program leaders pointed out that the cost savings may not be apparent in the short term but should materialize in the long term. The up-front costs include AV fistula surgery, working with the patients, and training. The savings should result from increasing transplantation and home dialysis, initiating dialysis without a hospitalization, and avoiding catheters and their associated risks and complications. They calculated that if all potential ESRD patients were to receive care using the Healthy Transitions approach, Medicare would save $2.4 billion a year, including savings from increased use of AV fistulas instead of catheters ($620 million), increased outpatient dialysis starts ($414 million), increased preemptive transplants ($825 million), and increased peritoneal dialysis ($550 million).

One nephrologist observed an improvement in patient education on nutrition and fluid management, as well as overall positive outcomes. The nurse care managers agreed that they were managing fluids and reported confidence that they were reducing the number of patients being admitted to the hospital for fluid overload because the patients were calling overnight and the nurses were able to give them advice before the patients proceeded to the ED.

In an interview, a nurse care manager said that Healthy Transitions was changing the way that CKD treatment is applied. One indication of this practice change was that a nephrologist stated he initially did not think that home dialysis without hospitalization was feasible, but was now having patients take this approach relatively often. Further, this nephrologist found that there were more timely referrals to the vascular surgeon because of the nurses' efforts to educate patients and their families.

Program leaders have evidence in their internal reports about increases in the number of "safe starts" (non-emergent initiation of dialysis) and a reduction in the use of central venous catheters, which in turn should reduce hospitalizations. In addition, the peritoneal dialysis rate was 29 percent in Healthy Transitions, almost three times the national average.

One aspect of the model that reduced costs was the medication reconciliation that the nurses conducted during home visits. This process reduced costs and improved patients' health status by eliminating unnecessary or duplicative drugs. Another cost reduction strategy was the

management of high-risk patients. One nephrologist related the story of a patient who had been hospitalized three times for heart failure. Since the Healthy Transitions intervention, the patient had not been hospitalized for the past year.

At least one-third of survey respondents identified the following factors as helpful in achieving program goals: shared understanding, participant engagement, and staff and clinician buy-in. None of the respondents reported that sufficient staffing or manageable documentation were helpful factors. In answering a question as to whether they have successfully engaged participants, almost all respondents strongly or somewhat agreed.

In a survey conducted by the implementation and monitoring contractor, program leaders cited the following as the biggest improvements they achieved: (1) increasing the preemptive transplantation rate (2 percent nationally versus 10 percent in Healthy Transitions) and (2) increasing the peritoneal dialysis rate to 29 percent (the national rate was 11.5 percent in 2013[7]).

In its final quarterly report, program leaders cited the following lessons learned: (1) nurse care managers needed to approach nephrologists differently for patients with eGFRs between 20 ml/min and 30 ml/min than they did for those below 20 ml/min because physicians were hesitant to refer patients to Healthy Transitions when the rate exceeded 20 ml/min; (2) the advantages of collaborating with the social worker to discuss care goals with the patient; (3) the importance of developing a standardized protocol to approach physicians about the benefits of program referrals and to address patient-centered issues such as palliative care, transplantation, and care goals that did not focus on modality selection; and (4) the need to streamline the program's data system to align with the nurses' workflow and potential future needs.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

For Healthy Transitions, patients should be exposed to the intervention for at least six months before they progress to ESRD, but results should have been demonstrated as early as the first program year. As noted, the net cost savings for enrollees might only be demonstrated after the end of the cooperative agreement. For example, a kidney transplant is an expensive procedure with large short-term costs, but a successful transplant can prevent years of dialysis, which can cost more than $88,000 per year.[8]

In addition to the cost savings, the program was intended to reduce ED visits and hospitalizations, increase peritoneal dialysis and transplantation rates, and increase AV fistula implants while reducing the rate of catheterization. These results should have been fairly consistent across implementation sites. Even if Healthy Transitions did not meet its internal quality measure goals, the program was expected to substantially exceed national benchmark rates for these metrics.

[7] See https://www.usrds.org/2015/download/vol2_USRDS_ESRD_15.pdf.

[8] This figure is based on 2014 data. See https://www.usrds.org/2016/view/v2_11.aspx.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Northwell's Healthy Transitions program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Northwell

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	177[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	88[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,948
Likelihood of all-cause hospitalizations	677
MDE sample size requirement to detect 20% effect	
Total expenditures	487
Likelihood of all-cause hospitalizations	169
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	Serious concern; we may be not able to identify a strong comparison group
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Too few treatment beneficiaries for primary outcomes
Survey data for treatment group that will be analyzed	Staff survey
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We do not expect to conduct a rigorous analysis of the impact of Northwell's Healthy Transitions program. The number of Medicare and Medicaid beneficiaries in the program remains too small to detect even substantial effects on hospitalization or expenditures.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents our third summary of the baseline characteristics of the treatment group. For the purposes of this report, the Northwell treatment group consists of Medicare FFS beneficiaries who were enrolled in the awardee's program between November 17, 2014, and May 31, 2016. In presenting the baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS (both Parts A and B), with Medicare as the primary payer when the beneficiaries' eligibility for awardee-provided services began (that is, their enrollment date) and who met all evaluation criteria for a period of at least 90 days during the baseline year (the 365 days immediately before their enrollment). The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. Of the 348 persons in the finder file, 111 were Medicare beneficiaries. Fourteen of these beneficiaries were not enrolled in FFS in the 90 days prior to enrollment, which left 97 participants who were included in the analysis of baseline characteristics for this report.[9]

The demographic characteristics of program participants were similar to those of Medicare FFS beneficiaries nationally (Table III.2). Most participants were either 75 to 84 years old (33 percent) or 65 to 74 years old (30 percent). Most were male (54 percent) and white (66 percent). Fifteen percent of participants were dually eligible for Medicare and Medicaid, compared to a national rate of 18 percent.[10]

[9] The enrollment figure of 629 participants in Figure II.1 is based on data reported by the awardee to the implementation and monitoring contractor through August 31, 2017, whereas enrollment figures for this section are from a finder file furnished by the awardee to us on enrollment only through May 31, 2016. Furthermore, we restricted this finder file to participants who were Medicare FFS beneficiaries and had been so 90 days prior to enrollment in Northwell's program.

[10] For national average rates, see the report on Medicaid enrollment from the Medicare Payment Advisory Commission and the Medicaid and CHIP Payment and Access Commission at http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-section-4-dual-eligible-beneficiaries.pdf?sfvrsn=0.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Northwell's program through May 31, 2016

	All participants (N = 97)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	16	16
65 to 74	29	30
75 to 84	32	33
85 and older	20	21
Gender		
Female	45	46
Male	52	54
Race		
White	64	66
Black	20	21
American Indian, Alaska Native, Asian/Pacific Island American, or other	9	9
Hispanic	3	3
Original reason for Medicare eligibility		
Old age and survivor's insurance	65	67
Disability insurance benefits	28	29
ESRD[a]	4	4
Hospice[b]		
Medicare/Medicaid dual status, percent dual[b]	15	15
HCC score[c]		Statistic
Mean		2.66
25th percentile		1.65
Median		2.37
75th percentile		3.33

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the initial intake visit by the RN care manager takes place. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

ESRD = end-stage renal disease; HCC = hierarchical condition categories.

Despite the demographic similarity of Northwell participants to the overall Medicare population, their utilization and care needs appear to be substantially greater than average. The mean hierarchical condition category (HCC) risk score of participants (2.66) is more than 2.5 times higher than the average for Medicare FFS beneficiaries nationally (approximately 1.00). Nearly all of the participants have HCC risk scores higher than the national average.

Participants had high and rising rates of service use and Medicare expenditures in the baseline year. The baseline utilization and expenditure data for a common set of measures are shown in Table III.3. We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $2,353—far above the U.S. average of $792.[11] Average PBPM Medicare payments for inpatient ($1,043) and physician ($665) services were the largest drivers of the total cost of care. Since the last report, these mean estimates have gone down, suggesting that recent participants are less costly and healthier.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Northwell's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	97	91	93	97	97
Average Medicare expenditures PBPM[a]					
Total	2,353 (389)	2,118 (459)	2,652 (646)	1,685 (418)	2,944 (506)
Acute inpatient	1,043 (291)	770 (240)	1,236 (465)	625 (309)	1,526 (366)
Inpatient other[b]	90 (48)	255 (169)	116 (115)	0 (0)	0 (0)
Outpatient[c]	173 (25)	184 (42)	142 (31)	184 (50)	181 (30)
Physician services	665 (60)	590 (59)	610 (65)	577 (69)	873 (85)
Home health	107 (22)	72 (35)	141 (46)	90 (34)	124 (37)
Skilled nursing facility	206 (76)	186 (131)	346 (173)	118 (78)	178 (88)
Hospice	9 (9)	0 (0)	0 (0)	37 (36)	0 (0)

[11] For national average rates, see the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation" at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed May 1, 2016.

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Durable medical equipment	59 (17)	61 (20)	61 (17)	55 (18)	61 (17)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	940 (198)	484 (151)	963 (271)	674 (227)	1,608 (314)
Outpatient ED visits	449 (84)	440 (146)	569 (146)	337 (117)	454 (142)
Observation stays	32 (18)	88 (62)	0 (0)	42 (41)	0 (0)
Primary care visits in any setting	8,226 (1,334)	6,066 (751)	8,094 (1,376)	7,957 (1,771)	10,639 (1,542)
Primary care visits in ambulatory settings	5,609 (527)	4,835 (620)	5,731 (638)	5,263 (728)	6,557 (842)
Specialist visits in any setting	29,582 (2,389)	26,462 (2,849)	27,519 (2,632)	25,218 (2,385)	38,722 (3,543)
Specialist visits in ambulatory settings	22,328 (1,532)	20,176 (1,852)	21,481 (1,869)	20,503 (1,682)	26,928 (1,817)
Measures of any health care utilization					
Percentage with a hospital admission[d]	52 (5)	11 (3)	16 (4)	15 (4)	28 (5)
Percentage with an outpatient ED visit[e]	34 (5)	10 (3)	14 (4)	8 (3)	10 (3)
Percentage with an observation stay[f]	3 (2)	2 (2)	0 (0)	1 (1)	0 (0)
Percentage with a 30-day readmission among all discharges	19 (4)	17 (17)	14 (7)	25 (11)	18 (7)
Percentage of participants with a readmission among all participants	8 (3)	1 (1)	3 (2)	3 (2)	3 (2)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

Table III.3 *(continued)*

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; PBPM = per beneficiary per month.

The rate of acute care hospitalizations for participants was 940 per 1,000 Medicare FFS beneficiaries per year during the baseline year—a rate much higher than the U.S. average of 274 per 1,000 Medicare FFS beneficiaries per year. Fifty-two percent of participants had at least one hospitalization during the baseline year. The rate of acute care hospitalizations for participants was highest in baseline quarter 4 (1,608 per 1,000 Medicare FFS beneficiaries per year) compared with baseline quarters 1 through 3 (484 to 674 per 1,000 Medicare FFS beneficiaries per year). About 28 percent of participants had at least one hospitalization during baseline quarter 4.

The rate of ED visits that did not lead to a hospitalization in the baseline year was 449 per 1,000 Medicare FFS beneficiaries per year. The rate of observation stays in the baseline year was 32 per 1,000 Medicare FFS beneficiaries per year. The rate of primary care visits (in any setting) was 8,226 per 1,000 Medicare FFS beneficiaries per year. The rate of specialty visits (in any setting) was 29,582 per 1,000 Medicare FFS beneficiaries per year. All of these rates were higher in baseline quarter 4 than in quarters 1 through 3.

In the baseline year, 19 percent of hospital discharges were followed by a readmission in the 30-day post-discharge window, and 8 percent of Medicare FFS beneficiaries had a hospitalization followed by a readmission in the 30-day post-discharge window.

In the near future, we will expand our reporting of baseline utilization and expenditure characteristics. Subject to receiving data from the awardee, we will, in future reports, include a descriptive trend analysis of how the intervention was implemented, including awardee-specific measures.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Northwell's proposed payment model combines a condition-specific, population-based payment for people with CKD (pre-ESRD) with a small set of quality-based incentive or penalty payments for nephrologists.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Northwell's proposed payment model is a population-based payment for people with CKD. Designed as a monthly care coordination and care management fee (the amount was not specified), the payment is intended to cover education and dietary assessments. In addition, the payment model includes a small set of quality-driven incentive payments to nephrologists that are designed to encourage better care, including earlier selection of dialysis modality (that is, before the disease progresses to the point of needing emergent vascular access for dialysis), as well as payment penalties for not achieving desired improvements. The incentive payments would reward nephrologists for AV fistula or graft placements, transplant listings, preemptive transplantation, initiation of home or outpatient hemodialysis or peritoneal dialysis, and conservative care management. Providers would be penalized for catheter placement, lack of hepatitis B and influenza vaccination, and anemia among patients who start hemodialysis.

C. Status of the payment model

The payment model was developed in collaboration with the National Kidney Foundation. Medicare is the primary payer of ESRD care and the most likely target audience for this program (even though it also includes pre-ESRD patients). The awardee has had preliminary conversations with some commercial managed care companies about participating in the proposed payment model. The awardee was considering discussions with New York Medicaid but had not met with Medicaid at the time of our interview.

D. Factors associated with the development of the payment model

Awardee leaders stressed that the incentives were generally not well aligned for providers or payers for this patient population. Because patients only become eligible for Medicare after progressing to ESRD, other payers see little return to any investments in the health of individuals with advanced kidney disease; the returns instead accrue to Medicare. The FFS system also provides incentives to dialyze patients in centers rather than with more cost-effective modalities, such as peritoneal dialysis. Thus, although the awardee's payment model attempted to realign incentives between payers and providers, awardee leaders suggested that the disconnect between the timing of investments and when savings occur would continue to be a barrier to a better payment model. They said that advice received from an expert consultant was helpful to their model development process, as well as collaboration with the National Kidney Foundation.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

As of the third program year, Northwell continued looking for sources of external funding to sustain its program. Northwell did not report activities to scale its program during the third program year. The awardee held conversations with a few Medicare ESCOs about replicating Healthy Transitions, but as of the end of the cooperative agreement none of these potential initiatives were finalized.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Northwell was considering different sources of funding to sustain its program after the end of the cooperative agreement. Awardee leaders reported that Northwell Health corporate leaders agreed to partially fund the program during the cooperative agreement and they were hopeful that the internal funding would continue after the cooperative agreement. In terms of funding from external sources, Northwell was working with the National Kidney Foundation to develop a payment model. The leaders were hopeful that Congress might approve a global capitation rate for dialysis in the near future.

C. Implementing the SSR plan: progress and changes

Sustainability. In the third program year, Northwell continued to consider different funding sources to sustain the program, but had not yet identified one. Awardee leaders continued to discuss the possibilities of using internal funding to sustain some program services and considered additional external funding options such as (1) generating revenue through the chronic care management codes and (2) offering the program through Northwell's managed care division as a product to some commercial payers as part of Northwell AdvantageCare. The awardee also held a meeting between several managed care companies and the National Kidney Foundation in July 2017 to discuss payment for kidney care management services. "We are in a period of uncertainty," one program leader said. "The future of the program is up in the air. Every possibility is open, including closing down the Healthy Transitions program."

Scalability. There were no efforts to scale the program during the third program year. The awardee reported that scaling was not feasible without additional resources.

Replicability. Northwell leaders have had preliminary discussions about replicating the program. Awardee leaders have spoken with the Medicare ESCO Rogosin about implementing the Healthy Transitions model and with Dialysis Clinic Inc. (DCI), which operates several ESCOs nationwide. Davita, a major dialysis provider that also participates in the ESCO model at multiple locations, expressed interest in the Healthy Transitions approach. Northwell reported potential plans to discuss its model with New York Medicaid but had not done so by the end of the cooperative agreement.

D. Factors associated with progress toward implementing the SSR plan

Northwell reported that payment policies and financial incentives challenged program sustainment, along with the political environment of the past year. First, dialysis facilities have a disincentive for patients to receive transplants, so Healthy Transitions' goal of increasing the number of transplants may be seen as a barrier. Similarly, nephrologists who treat patients on dialysis would "lose" patients who receive a transplant. Second, political uncertainty about the fate of the Affordable Care Act of 2010 during the last program year made it difficult to engage payers, according to awardee leaders.

Northwell also decided to wait to approach payers about funding the program for Medicaid patients. The program would likely have to be adapted to better serve Medicaid patients, who can be more complex and difficult to treat and follow up with, according to a Northwell leader. "Medicaid patients by nature of the insurance are sometimes very difficult to move [in terms of their health outcomes]. It's hard to even get them a connection with a surgeon in order to place access for renal replacement therapies," an interview respondent explained.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. However, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or the sustainability, scalability, or replicability plans. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: New York City Health + Hospitals

April 18, 2018

Stefanie Pietras (Mathematica Policy Research)
Rivka Weiser (Mathematica Policy Research)
Eric Lammers (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

This page has been left blank for double-sided copying.

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

New York City Health + Hospitals, a public benefit corporation that serves as the public safety net in the city's health care system, used funding from HCIA R2 to develop and support the ED Care Management Initiative at 6 of its 11 hospitals. The program, which is based on a pilot program by the awardee, provides care management to participants while they are in the ED

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

along with 90 days of follow-up care coordination to help participants better manage their health and avoid unnecessary ED visits and hospitalizations.

Care management is provided by nurses, who generally use data in electronic medical records (EMRs) or talk to on-duty ED clinicians to identify potential participants while they are still in the ED. After recruiting and enrolling patients in the program, the nurse care managers (NCMs) provide initial care management services by doing a risk assessment, creating an ambulatory care plan, providing health education and counseling, and coordinating referrals to other providers.

The NCMs link participants to other care team members, such as a home care intake nurse (for an initial assessment of home care needs and to arrange home care if appropriate) or a pharmacist (for medication management or other services). Care management staff try to schedule a follow-up primary care (or specialist) visit for the participant as appropriate, and the NCM checks in with the participant by phone within 24 to 48 hours after discharge from the ED. Thereafter, community liaison workers (CLWs) provide transitional care coordination via phone. For example, the CLWs may remind participants of upcoming ambulatory care visits and follow up with them after these visits. They may also link participants to other providers and resources, and check in with participants at 30, 60, and 90 days post-enrollment (the intervention ends at 90 days). The program is tailored to local needs and capacity, so the roles and responsibilities may vary somewhat from one hospital to another. Some tasks typically performed by CLWs may be assumed by NCMs, and vice versa.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The program provides care management and 90-day care coordination (which covers the transition to comprehensive ambulatory care) to eligible patients who visit the ED.
Major innovation	An ED-based team of NCMs, pharmacists, community liaison workers, and home care intake nurses help patients to better manage their health by providing education, support, and linkages to ambulatory care and home health care as needed.
Program components	• Care management • Transitional care coordination
Target population	Adult patients are eligible for the program if they meet any of the following requirements: (1) Visit the ED for ambulatory care sensitive conditions (ACSCs) (2) Visit the ED and meet particular utilization-based criteria (3) Are deemed likely to benefit from the program based on the clinical judgment of the nurse care manager or referring ED clinician
Theory of change/ theory of action	New York City Health + Hospitals hypothesized that providing interdisciplinary care management and extended care coordination would help ED patients with ACSCs and other patients whose conditions are high-risk/high-cost to better manage their health and avoid unnecessary hospitalizations and repeated visits to the ED, which would lower the cost of their care.
Payment model	Global risk capitated contracts; value-based payment approach
Award amount	$17,916,663

Table I.1 *(continued)*

Program characteristic	Description
Effective launch date	9/1/2014
Program setting	ED with follow-up via telephone
Market area	Urban
Market location	New York City
Target outcomes	• Reduce 30-day hospitalizations by 35 percent • Reduce 7-day and 30-day repeat ED visits by 25 percent • Produce net savings of $75 million

ED = emergency department; NCM = nurse care manager.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was partly successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on three factors.

First, the awardee partially achieved enrollment effectiveness, enrolling 83,946 participants—83.8 percent of its enrollment target—by the end of the initial cooperative agreement. Second, pressure to increase enrollment, demanding workloads, staffing vacancies, and limited capacity of primary care providers challenged the awardee's ability to provide comprehensive care management, provide all of the expected follow-ups, and secure timely ambulatory care appointments. Third, although patients were relatively easy to engage in the ED, it was often challenging to reach and engage them by phone after they were discharged. Despite these challenges, participating clinicians and other implementation staff reported that the program had a positive effect on the delivery of care.

Impact evaluation. We find no evidence that the New York City Health + Hospitals program has lowered total spending by Medicare, cut down on the participants' hospitalizations, or helped participants stay out of the ED among Medicare fee for service (FFS) beneficiaries. In fact, the percentage of Medicare FFS treatment beneficiaries with an ED visit increased 2.2 percentage points above what would have been predicted if there had been no program. The number of visits by beneficiaries to primary care providers and specialists decreased, along with payments to physicians.

Payment model. Instead of developing a stand-alone payment model for its program services, New York City Health + Hospitals is incorporating its program into its existing contracting vehicles. These vehicles include global risk capitated contracts with Medicaid and Medicare managed care plans and a value-based payment approach that aligns with broader New York Medicaid reform efforts. The awardee is also in the early stages of exploring shared savings through an all-payer accountable care organization (ACO).

Sustainability plans. New York City Health + Hospitals plans to sustain and scale its program after the cooperative agreement by implementing a similar Delivery System Reform Incentive Payment (DSRIP)–funded program that builds on the HCIA R2–funded program. The

sustained program will be different in some ways, with changes based on recommendations and feedback from internal stakeholders. The awardee implemented the HCIA R2–funded program at six of its hospitals, and plans to scale the program to all 11 hospitals after the cooperative agreement ends. The awardee did not report plans to replicate its program.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of non-clinicians on their perceptions of the effect of the programs on the delivery of care. That survey was fielded around the start of the third program year with a sample of 47 potential respondents and achieved a response rate of 72 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partly successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

NCMs used information from sources such as EMRs, electronic white boards, and conversations with ED clinicians to identify potential participants. Each hospital site also had a physician advisor for the ED Care Management Initiative, who educated ED clinicians about the program and encouraged ED clinicians to refer patients to NCMs. However, ED clinicians had varying levels of engagement with the program, depending on their hospital site.

Patients were eligible for the program if they had visited the ED with an ACSC or met certain utilization criteria (for example, they were hospitalized or visited the ED within the past 30 days). Even if they did not meet these criteria, patients who visited the ED could be eligible for the program if a referring ED clinician or NCM thought they could benefit from it.

In January 2017, the awardee reduced its monthly enrollment targets by 30 percent in order to address issues with staff workload (discussed below). At that time, leaders and physician advisors also revisited the eligibility criteria. The group agreed that the criteria should be more flexible and allow for greater clinical judgment to identify those who were expected to benefit most from the program, such as high-acuity patients, high utilizers, and those with social and behavioral health issues. Program leaders and staff also acknowledged that some patients who met the official criteria, such as those with urinary tract infections (UTIs) (an ACSC), did not need the comprehensive care management and care coordination the program offered. Physician advisors, in turn, carried back the message to sites that NCMs could use more clinical judgment in prioritizing patients whom they thought would benefit from the program. The official criteria remained the same in order to keep a consistent enrollment approach across the three years, despite the greater reliance on clinical judgment toward the end.

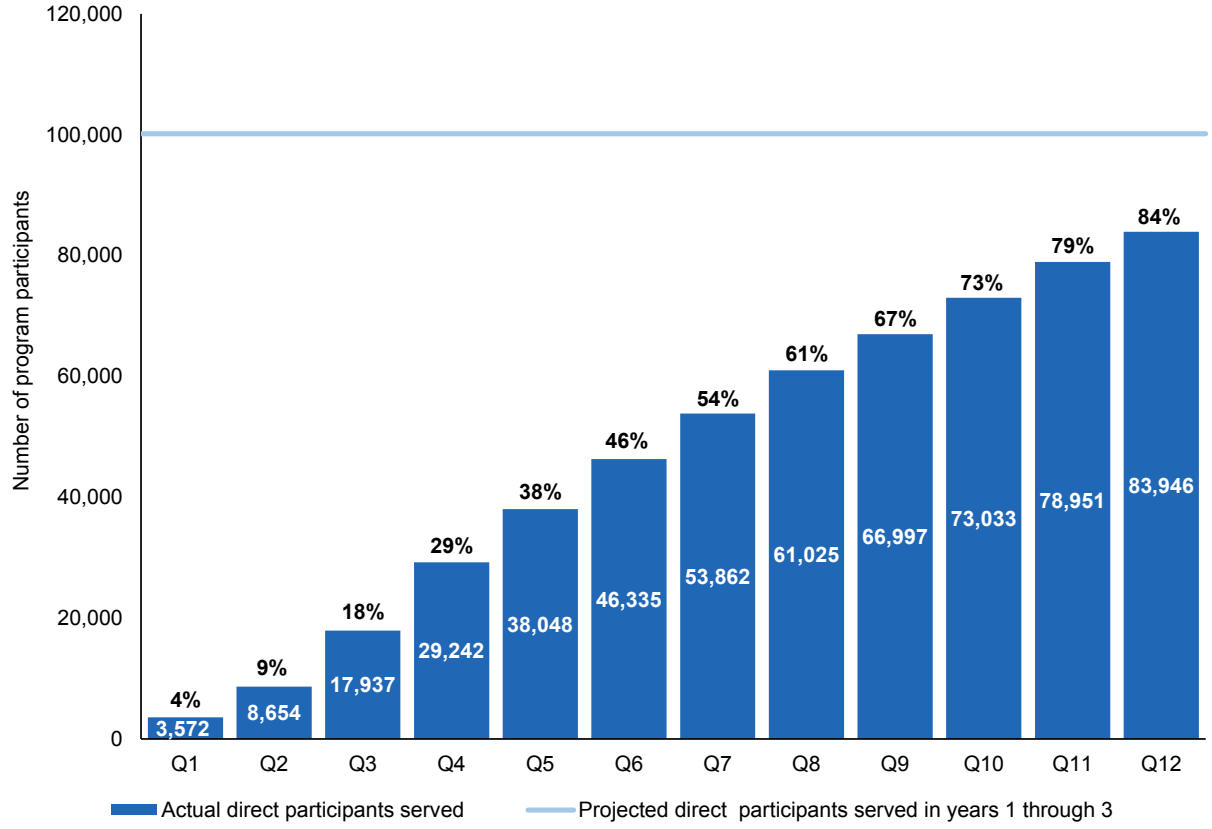
b. Evidence of enrollment effectiveness

New York City Health + Hospitals partially achieved enrollment effectiveness. Overall, the awardee reported that it enrolled 83,946 unique direct participants from September 2014 (when it launched its program) through August 2017. This represents about 84 percent of its final three-year projections (Figure II.1).

Despite not meeting its final enrollment target, enrollment in the New York City Health + Hospitals picked up over time, even exceeding monthly enrollment targets in many months. Enrollment was slow in the beginning of the first year, when the program was not fully staffed due to bureaucratic hiring processes and a competitive job market.

In March 2015, CMMI asked the awardee to develop a plan to address its slow implementation and low enrollment. Program and site leaders told staff to place a high priority on increasing enrollment. This emphasis, along with the awardee's achievement of full staffing by the end of the first year, enabled New York City Health + Hospitals to nearly meet its first year's enrollment target (89 percent). For most of the program period, meeting monthly enrollment targets was a top priority for staff, and the results of this dedication can be seen in the awardee's self-monitoring data. The awardee had a seven-month stretch in 2015 during which it consistently exceeded its monthly enrollment targets.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding.

c. Barriers and facilitators associated with enrollment effectiveness

New York City Health + Hospitals' progress in meeting its three-year enrollment goals was influenced by several factors. As noted, there was a push early in the program to focus on enrolling patients. NCMs used various strategies in an attempt to meet their targets, such as recruiting in the ED's urgent care area because patients there were more likely to meet the program criteria. However, the focus on enrollment placed stress on the NCMs and CLWs as they tried to cope with their workloads. Therefore, leaders sought CMMI's permission to reduce enrollment targets in the final year. Leaders decided not to share the exact revised targets with staff in order to reduce the emphasis on enrollment numbers. Leaders reported that NCMs expressed appreciation for the lower numbers, but interviews with physician

> "After we got the approval from CMS, we really decided that as part of the justification for asking for the decrease in overall enrollment in the third year that we really needed to take a look at the patients that we were actually seeing because we're noticing that … a lot of times our facilities were just going in and grabbing all the patients that met the criteria, when the honest truth was that people with a UTI don't necessarily need 90 days of care management."
>
> —Project leader

advisors and frontline staff revealed that the sites were still diligently focused on recruitment: "We still had a sense of attack, even without the official numbers being out there; we still just continued working and did what we could to meet our goals and our target," one NCM explained. Frontline staff, however, shared examples of their efforts to broaden their target population. One NCM noted that the ED director had worked with the NCM's team to widen the scope of who was enrolled at that site to target those who might otherwise "fall through the gaps." Leaders reported they did not see any major changes in enrollment numbers or in the types of patients who enrolled, but they did acknowledge that staff were likely already conditioned to their enrollment approach by the program's final year.

Clinician engagement was either a barrier or facilitator to enrollment efforts over the course of the cooperative agreement, depending on the site. Resident turnover made it hard to keep clinicians engaged. Sites tried various strategies to overcome this, with mixed results. Strategies included creating mousepads with pictures of NCMs, posting NCMs' schedules in ED workstations (sometimes with pictures of and information about NCMs), and having the NCMs approach providers to introduce themselves. Site leads and physician advisors helped to spearhead some of these efforts.

Although some sites reported that these strategies did not make a difference, other sites had success with them over time. One NCM spoke highly of her team's collaborative relationships with clinicians and attributed her site's success to having dedicated workspace and cellphones, so that NCMs were easy to reach. Physician advisors interviewed in the third year explained that clinicians at their sites who were originally skeptical of the program had come to appreciate it, giving examples of times when clinicians were disappointed to find NCMs were unavailable and when clinicians sought out NCMs to help when particular patients returned to the ED.

Technology also both facilitated and hindered enrollment efforts. EMRs and electronic whiteboards helped alert NCMs when patients met diagnosis or utilization-based eligibility criteria, but it was harder to find other information to help determine whether a patient was appropriate for the program. NCMs needed to interact closely with ED clinicians in order to determine whether a patient would likely be admitted, as admitted patients were not eligible for the program. The Access care management database that was used by the sites was not interoperable, so NCMs were unable to check whether a patient was recently seen or enrolled at another New York City Health + Hospitals site. In the third year, leaders noted that the program would switch to GSI, an interoperable care management software, once the program is sustained under DSRIP. Details on the program's transition to DSRIP may be found in Chapter V.

Finally, when patients were approached, they were generally receptive to the program, and this facilitated enrollment. Patients who were not interested were often focused on getting their immediate needs taken care of by an ED physician after a long wait in the ED.

2. Delivery of program services

a. Description of and changes to service delivery model

The ED Care Management Initiative built on a pilot program that began in 2008 at two New York City Health + Hospitals sites and was designed to address overutilization of the ED. In 2011, the program expanded to all 11 hospitals in the system. Each hospital had one or two

NCMs, who provided care management and linked patients with ACSCs to primary care. In developing the HCIA R2–funded program, New York City Health + Hospitals expanded on its earlier ED care management efforts by including an extended 90-day follow-up, adding pharmacists and CLWs to the care management team, increasing resources for care management, and enhancing standardization across sites.

In the ED Care Management Initiative, NCMs recruited participants based on the eligibility criteria in Table I.1. After patients gave their informed consent, NCMs performed a risk assessment, created an ambulatory care plan, provided health education and counseling as needed, referred participants to primary care and/or specialists as appropriate, and referred appropriate participants to the home care intake nurse and pharmacist. Generally, the NCM called the patient to check in about 24 to 48 hours after discharge.

Upon referral, the home care intake nurse assessed the participant's need for home care and linked him or her to appropriate services. The pharmacist provided various services based on participant needs, such as educating the participants about medications or helping them access necessary medications.

The CLW (and sometimes NCMs, depending on the site) reminded participants about their appointments and followed up with them after appointments; asked participants if they had questions and linked them as needed with NCMs and pharmacists for more support; and checked in on participants throughout the 90 days of program enrollment. Based on these efforts, the awardee theorized that participants' health conditions would be managed better than they were when the participant had the qualifying ED visit, because their ability to self-manage care would improve and they would engage with appropriate care. In turn, this enhanced management would lead to fewer return visits to the ED and fewer hospitalizations and generate a net savings of $75 million.

b. Evidence of service delivery effectiveness

Overall, New York City Health + Hospitals partially achieved service delivery effectiveness. Pressure to increase enrollment, demanding workloads, staffing vacancies, and primary care capacity issues challenged the awardee's ability to provide comprehensive care management, provide all of the expected follow-ups, and secure timely ambulatory care appointments. Further, although patients were relatively easy to engage in the ED, it was often challenging to engage them by phone after they were discharged. We provide details below.

Delivery of intervention services. In our interviews, frontline staff noted that a heavy emphasis on enrollment meant the NCMs had less time to provide comprehensive care management; staff thought the quality of their services suffered as a result. In addition, the original staffing structure of one CLW to five NCMs at each site may have limited the delivery of the intervention. Respondents noted that, instead of the follow-up calls at 30, 60, and 90 days in the original intervention design, some sites completed only certain calls or prioritized calls for participants with certain conditions due to staffing and time constraints. The awardee did not collect centralized data on the number of follow-up calls it successfully completed.

Linking participants to ambulatory care was a key component of the program and an ongoing challenge due to limited capacity of primary care providers. The average time elapsed

from the ED visit that precipitated enrollment to a primary care appointment ranged from 22 to 35 days in the period from September 2014 to April 2017; the awardee was never able to meet its target of 14 days or less during this time frame, and there was no sustained improvement over time. Similarly, the proportion of participants who saw a New York City Health + Hospitals primary care provider within 90 days of their initial ED visit decreased from 37 to 27 percent between September 2014 to April 2017; the awardee never met its target of 60 percent or higher.

Although these measures had limitations (such as not capturing appointments with community providers) and are accordingly underestimates, they demonstrate the challenge of facilitating prompt primary care appointments within the New York City Health + Hospitals system. Also, as physician advisors pointed out, these metrics did not capture efforts to link participants to specialists, because some participants were referred to appropriate specialists instead of primary care providers. Participants were included in the denominator of the measure assessing the proportion of participants who saw a primary care provider within 90 days even if they did not receive a referral to primary care.

The awardee's self-monitoring data indicate that home care services were used throughout the intervention, with a dramatic dip in referrals midcourse. From September 2014 to March 2016, the number of home care referrals increased from 24 to 80. After this period of progress, referrals dropped to about two per month in summer 2016. Leaders and managers were alarmed by this stark dip in numbers and launched an intervention to revive the home care component. By spring 2017, the awardee averaged around 40 referrals a month.

Although the awardee did not share self-monitoring data on the pharmacist component, staff repeatedly emphasized the value of the pharmacist services. Pharmacists did not convey the same level of concern about workload in interviews as other staff members did. This intervention component seems likely to have been delivered as it was designed to be.

Staffing and training. Staffing was a major challenge for the awardee—both at the management and frontline levels. The awardee was not able to hire its data analyst, financial analyst, or project manager until the end of the first year. Further, the program was led by three different project directors and two different project managers over the course of the cooperative agreement.

The awardee also did not achieve full staffing at sites until the end of the first year. According to data shared by the awardee, the program typically lost a staff member every quarter. Filling these positions again proved to be challenging, and they remained vacant for long periods. Half of the non-clinicians surveyed believed staff turnover and unfilled positions were a barrier to the program's achieving its goals.

In the first year, CLWs received a 35-hour training, but NCMs and other staff received only minimal onboarding training. In the second year, the entire care management team participated in a multi-day training. In the beginning of the third year, 79 percent of non-clinicians who answered the survey agreed that the training had helped to improve their job performance, although 45 percent wished they had been offered other training that could have helped them meet their responsibilities within the program. In the third year, leaders organized a training on the social determinants of health.

Recruitment and engagement of providers. Implementation of the ED Care Management Initiative did not require the awardee to recruit or engage providers or provider organizations.

Engagement of program participants. In the staff survey, over 90 percent of non-clinicians agreed that the program had successfully engaged participants. However, 50 percent of staff noted that resistance to the program by participants was a barrier to achieving the program's goals. As noted, the frequency of follow-up varied by site, and some sites chose to prioritize particular participants for follow-up.

Further, in interviews CLWs consistently brought up the challenge of reaching and effectively engaging participants over the phone. It often took more than one phone call to reach a participant, and sites varied in the number of repeat calls that staff were willing to make given the demands on their time. In addition, staff thought participants sometimes gave incorrect contact information or did not answer the phone because they were nervous about legal, immigration, or financial issues. When participants did answer the phone, some of them did not remember being enrolled in the program. The low proportion of participants who kept their primary care appointments may also be an indicator of unsuccessful participant engagement, because they may not have been motivated enough to attend their appointment or may not have been reminded about it, among other reasons. One staff member in the third year talked about repeatedly scheduling appointments for participants who never showed up for them.

c. Barriers and facilitators associated with service delivery effectiveness

Establishing linkages with ambulatory care providers was an ongoing challenge for the awardee because those providers had limited capacity. Sites implemented various strategies to improve collaborations with ambulatory care. Physician advisors worked closely with ambulatory care in their facilities to develop site-specific care pathways (that is, to establish workflows across care settings for specific conditions).

One site, for example, worked with cardiology on a process to secure urgent cardiology appointments within 48 to 72 hours. In addition, based on care pathway work at sites and collaborations across settings, some care management staff knew exactly who to contact with particular questions or concerns. For example, some CLWs who had trouble securing an appointment for a participant reached out to an established contact in the administration of an ambulatory care clinic for help linking the participant to appropriate care. Although physician advisors often helped forge relationships with ambulatory care settings, care management teams also took the initiative. One care management team decided to partner with its hospital's primary care clinic in the third year to monitor the wait time for primary care

> "It's a lot about coordination of care. It's a lot of work because the institution has all these moving parts, and they have to kind of run at the same pace. Otherwise, the patient falls through the cracks. It's not easy. It's a kind of work that a lot of the practices are not used to and requires a lot of meetings. But, now they're more used to it. They know when to call me, when to email me, or go to the administrator instead of the physician. So … we are learning how to do this because it's a banner of ownership. Ambulatory care is learning to own their patients. We are here to help the patients go back, after troubleshooting the problem. So, this is a learning process. It's not just creating the mechanism for flow, but also the way people think. It's a culture shift. Ultimately … a culture shift [is] occurring. And that is an amazing thing, because it's so hard to do."
>
> —Physician advisor

appointments because at one point, there was close to a 30-day wait for appointments. Each week, the team sent a template to the medical clinic's administrator, showing the timing of appointments that were obtained. By reporting data weekly and contacting the clinic's administrator to overbook appointments as needed, the team was able to shorten the time spent waiting for appointments.

Another care management team took the initiative to forward participants' information to outpatient NCMs (who happened to be located in the same hallway) when they wanted to ensure a participant received certain services, such as nutrition classes for a diabetic participant. As physician advisors explained, integrating care across settings was a long-term vision, and the system was just beginning to establish workflows, relationships, and infrastructure to facilitate this transition. In the third year of the program, central leaders also encouraged sites to refer to Gotham sites, which were standalone clinics within the New York City Health + Hospitals system. Although some Gotham sites had limited capacity, this strategy reportedly improved linkages for some program sites.

The program also had some difficulty linking participants to home care in the beginning because many participants did not have an established primary care provider, and home care orders need to be approved by a physician. As a workaround, home care intake nurses asked ED clinicians to sign the initial order. The clinicians often were resistant because they did not want to be responsible for participants' care once they left the ED. Over time, home care intake nurses reported less resistance as physicians became more comfortable signing orders when they knew participants had appointments scheduled in the near future with primary care providers.

Linking participants to home care also was a challenge for the awardee, as evidenced by the dip in referrals midcourse. Home care intake nurses were only on the program half time, so they were not always in the ED. Some sites reportedly preferred to call and refer participants rather than engage home care intake nurses in more active discussions. To revive home care efforts, home care leaders encouraged the home care intake nurses to be more active members of teams and encouraged teams to include the nurses in more discussions and meetings. The nurse lead for home care intake reported that the home care intake nurses at multiple sites had been invited to attend morning rounds and regular meetings, so she thought these conversations had been effective.

Intense workloads were a recurring issue throughout the project, a concern recognized by leaders and staff alike. Two-thirds of non-clinician staff surveyed indicated that insufficient staff time for the amount of work was a barrier to the program achieving its goals, and nearly 40 percent indicated that the program increased their feelings of burnout. By the second year of the program, care management staff consistently revealed in interviews their concern that the program was focused on "quantity over quality." Staff hypothesized that the lack of sustained improvement in program outcomes could stem from this issue. Leaders also wondered if a more targeted intervention with greater intensity would lead to better outcomes, and accordingly reduced enrollment targets in January 2017. As noted, the reduction did not produce significant changes in ingrained staff behavior, but some staff said they were trying to prioritize higher risk participants who could most benefit from the program. In addition, as noted, bureaucratic hiring processes and a competitive job market contributed to slow initial hiring and difficulty back-filling positions throughout the cooperative agreement.

Leaders acknowledged that the original HCIA R2 application was written mainly by staff in finance and would have benefited from more clinical involvement, which would have better reflected the clinical resources needed to implement the intervention. The awardee reported feeling generally tied to what was written in the application for the cooperative agreement, but planned on using lessons learned from the experience to inform the future program under DSRIP.

Throughout the program period, some sites adapted the service delivery model in response to local needs and staffing limitations inherent in the staffing model (that is, one CLW per site). For example, at one site, NCMs stepped in to conduct the follow-up calls after the site lost its CLW; according to the NCM who was interviewed, the site found that nurses were better suited than non-clinical staff to answer follow-up questions, especially about medications. Over time, some sites increasingly targeted calls to those with more chronic conditions such as hypertension and diabetes, whereas others flagged participants who were especially sick for follow-up and prioritized calls accordingly. In the second year, the awardee used carryover funds to hire a second CLW at two sites; bureaucratic barriers interfered with the hiring of CLWs at the other sites. Some sites provided additional in-kind CLWs.

The awardee realized that the social determinants of health were critical to address and acknowledged that its intervention did not have robust resources available to identify and meet these needs. Frontline staff in interviews noted that some sites had social workers in the ED (not funded by the cooperative agreement) who often played a critical role in care management services and should be part of the official care management team. When the awardee applied a risk stratification methodology developed by its Office of Population Health to the enrolled population to define a "high-risk" group (defined as the top 1 percent of risk-scored participants), it found that blacks and Hispanics accounted for 75 percent of the high-risk group, and 18 percent of this group was affected by homelessness in 2016. To address social issues, leaders organized a training on the social determinants of health in the third year for care management staff. In the next phase of work under DSRIP (described in Chapter V), the care management teams will include a social worker. The DSRIP program will also enable CLWs to conduct home visits with higher-risk patients as an additional strategy to promote engagement and address social determinants of health.

Burdensome documentation and data systems were key challenges throughout the cooperative agreement. Over half of non-clinicians surveyed in the beginning of the third year thought that issues using program technology, including EMRs, were a barrier to the program achieving its goals. Even more (69 percent) thought documentation requirements were a barrier. To document participant information, care management teams generally completed paper forms, entered clinical information into EMRs, and entered program data into an Access care management database.

In the second year, two sites switched to the Epic EMR, which reduced duplication, because these sites did not have to enter information in the Access database as well. Although staff at these sites noted that Epic streamlined work, provided easier access to data, and enhanced communication, some still noted examples of duplicate documentation or times when multiple-choice options were not comprehensive enough. A cross-site workgroup helped to optimize the functionalities of Epic for the program. In the third year, a program leader indicated that

documentation continued to be a significant challenge, noting that data were not always being documented appropriately or in the right system(s). The documentation challenges meant the awardee was unable to centrally track key process measures (such as the number of follow-up calls completed).

Overall, staff found trainings to be beneficial, but some expressed interest in earlier trainings or other types of training. In the first year, program leaders commented that CLWs appreciated the 35-hour training and found it valuable. In the second year, frontline staff reported that the multi-day training was helpful overall, particularly with improving their communication with participants. Some wished they had received the training earlier, and others commented that time pressures prevented them from fully applying what they learned. Some staff reported that they found the Year 3 training redundant with earlier trainings and considered it a refresher course. NCMs and pharmacists also expressed their desire for other types of training in Year 3 interviews, noting a particular need for training to become a certified diabetes educator and a certified asthma educator.

Program leaders' emphasis on cross-site learning and quality improvement also facilitated service delivery. The project manager praised an earlier project director's decision to establish workgroups from executive to facility levels which involved the physician advisors, pharmacists, and site leads. She noted that the workgroups helped to spread best practices, develop consensus, and increase the visibility and transparency of the work being done. In addition, workgroup meetings served as brainstorming sessions, which helped to inform widespread program changes and to inform the program as it will operate under DSRIP.

Workgroups and individual sites used program data to inform improvement efforts. In interviews, frontline staff often referenced site-specific reports developed by the central office, which helped them to develop strategies for improvement at their site. In the third program year, the project manager also conducted a site visit to each site to observe workflows, handoffs, and documentation and to discuss program implementation with staff. The site visits helped leaders select areas to focus on for the remainder of the cooperative agreement: improving participant linkages to primary care and home care, and moderating the strong emphasis on enrollment.

C. Assessment of perceived program effects on the delivery of care and outcomes

In Year 3 interviews, respondents thought that the program's core goals (reductions in ED visits, hospitalizations, and costs) were attainable, though most interviewees qualified their responses by suggesting that the program had limitations, which would reduce its impact. A few staff referenced the awardee's initial analyses, which found that the program significantly reduced repeat visits to the ED, but not hospital admissions, within 30 days of enrollment. These staff thought those findings made sense, explaining that the return ED visits were likely avoided because (1) participants were educated on the appropriate use of the ED, (2) pharmacists and NCMs had coached participants on managing their care, and (3) participants had been successfully linked to ambulatory care and home care. This meant less use of the ED for unnecessary reasons, but these staff expected that the intervention might not impact hospital admission for the sickest participants. One program leader was confident that the program would achieve cost savings, but thought it was more challenging for the program to significantly impact

other outcomes because of the wide variation between facilities and the ongoing struggle with linking participants to primary care. Physician advisors noted that they thought eligibility criteria were too broad and the program likely would only impact certain patient populations (for example, the higher-risk patients). They also thought that this cooperative agreement helped to provide the impetus for system change, and efforts made across New York City Health + Hospitals were only just beginning to build relationships and infrastructure to coordinate care across settings, address the social determinants of health, and anchor patients' care in the ambulatory care setting.

Non-clinician staff survey results from the beginning of the third year reinforced sentiments shared during the interviews. Over three-quarters of respondents believed the program had been somewhat or very effective in achieving its goals. Over 90 percent thought the program had a positive impact on the quality of care and services provided to participants; a positive impact on the efficiency of care or services provided to participants; and a positive impact on the ability to provide care or services responsive to participants' preferences, needs, and values. More than four in five also believed the program had a positive impact on the ability to respond in a timely way to meet participants' needs (85 percent), and that the program had a positive impact on the access to care and services for all patients (82 percent). Although the scores are still relatively high, these slightly lower scores might reflect staff's awareness of primary care capacity issues and possibly the effects of staffing and time constraints on providing prompt and comprehensive program services.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

Our assumptions are predicated on New York City Health + Hospitals' theory of action being accurate in terms of the planned processes resulting in the desired outcomes. The awardee reflected toward the end of the program that having a patient attend a primary care appointment may not do enough to avoid repeat ED visits by itself and that it is important to address the social determinants of health. Although New York City Health + Hospitals worked to address social factors as best it could, the awardee acknowledged that including social workers on its interdisciplinary teams would have been a big help.

Nearly all participants were likely to have been enrolled long enough to receive all intervention services. As noted, the intervention was perceived by some staff as likely to have a greater effect on repeat ED visits than on hospitalizations.

Some aspects of the awardee's eligibility criteria are difficult to apply to an impact design, particularly the category based on clinical judgment. However, there should be enough participants with ACSCs to analyze. As the awardee encouraged staff to target high-acuity patients and use greater clinical judgment beginning in January 2017, there may be slightly more participants in the clinical judgment category that will need to be excluded from analyses after this date because there will not be known criteria to match on. The program may also show greater impacts shortly after this date because staff were instructed to focus on providing higher

intensity care to high-acuity patients and decrease their emphasis on enrollment. Also, of note, the program may have more of an impact for patients with particular ACSCs or utilization criteria.

There will likely be differences in program impacts across sites, because sites implemented various adaptations to the service delivery model to meet local needs and address staffing constraints. Sites that were fully staffed for the duration of the program or had additional in-kind support staff may have better outcomes than sites that were challenged by staffing vacancies or only had one CLW. Methods of linking to primary care also likely have an effect on program outcomes. Some sites were more active in linking participants to primary care by making appointments and reminding participants about their appointments, while at other sites, participants had to be more proactive in arranging for primary care appointments.

Participant demographics also varied by site. The awardee reported that its high-risk participants were primarily concentrated in the Bronx and Queens (four of the six sites). If the awardee's theory that high-risk patients are more likely to benefit from the program is correct, program impacts may be stronger at hospitals in these locations.

Finally, as all of the New York City Health + Hospitals sites participated in the pilot program, and therefore have had some exposure to ED care management, using the non-HCIA R2 sites as matched comparisons may underestimate the effects of the program.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter updates our assessment of the evaluability of the New York City Health + Hospitals program, baseline characteristics of the treatment group, our comparison group matching procedures and baseline balance statistics, and preliminary impact findings for Medicare fee-for-service (FFS) beneficiaries.

Key findings

- We find no evidence that the NYC H+H program has lowered total spending by Medicare, hospitalizations, or ED visits among Medicare FFS beneficiaries.
- The percentage of treatment Medicare FFS beneficiaries with an ED visit was 2.2 percentage points greater than the 42 percent that we estimate would have had an ED visit if there had been no program.
- The number of visits to primary care providers and specialists was 10 to 13 percent lower than the numbers we estimated had there been no program; similarly Medicare expenditures for physician services were 10 percent lower than the amount we estimated would have been spent had there been no program.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

New York City Health + Hospitals' program ended August 31, 2017. The projected sample sizes in Table III.1 are based upon actual enrollment as of February 28, 2017 and are the number of beneficiaries that we estimate will be included in our evaluation to allow for all participants to receive up to six months of program exposure (the 90 day period of the care management program plus an additional 3 month follow-up period during which the participant may continue to receive other ambulatory care if delays in care were encountered during the initial 90-day period). Full program exposure is a requirement for inclusion in our evaluation to minimize dilution of program effects due to limited exposure to the program components. The projected numbers also reflect application of eligibility criteria for inclusion in our evaluation, such as Medicare FFS versus Medicare managed care. Due to processing lags in Medicaid enrollment data, we are projecting that 6,907 Medicaid beneficiaries meet program eligibility for inclusion in our impact evaluation but there may be a substantially larger number that we can include once we receive more timely enrollment data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: New York City Health + Hospitals

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure as of February 28, 2017	3,015[a]
Projected Medicaid population with 6 months of program exposure as of February 28, 2017	6,907[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	3,021
Likelihood of all-cause hospitalizations	1,361
MDE sample size requirement to detect 20% effect	
Total expenditures	755
Likelihood of all-cause hospitalizations	341
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	Some issues, but probably surmountable. Expect to select a comparison group
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We will be conducting a rigorous impact evaluation of the awardee's program by using difference-in-differences estimation with propensity score–matched comparison groups. Separate analyses will be conducted for Medicare and Medicaid adults (age 18 and older). The comparison groups will be comprised of those who have an ED visit to one of nine other hospitals in New York City, including the five EDs in the awardee's system that did not have the HCIA R2–funded program. We have developed a comparison group for Medicare FFS beneficiaries who met these eligibility criteria between program startup and November 30, 2016 and are reporting interim impact findings in this report. For the final report, we will develop a comparison group of Medicaid beneficiaries should T-MSIS data files become available. We estimate that enough Medicare and Medicaid beneficiaries will meet our eligibility criteria to detect a 10 percent effect or larger on Medicare and Medicaid expenditures.

B. Characteristics of Medicare and Medicaid participants at baseline

This summary covers both the common and awardee-specific claims-based outcomes at baseline for Medicare FFS beneficiaries in the treatment group who enrolled in the awardee's program from September 2014 through November 2016, according to lists from the awardee. To be eligible for the program, a patient who visits the ED must be able to be discharged from the ED safely and either have (1) visited the ED for an ambulatory care sensitive condition (ACSC); or (2) met particular utilization-based criteria (for example, had another recent ED visit or hospitalization); or (3) been deemed likely to benefit from the program based on the clinical judgment of the NCM or referring ED clinician.

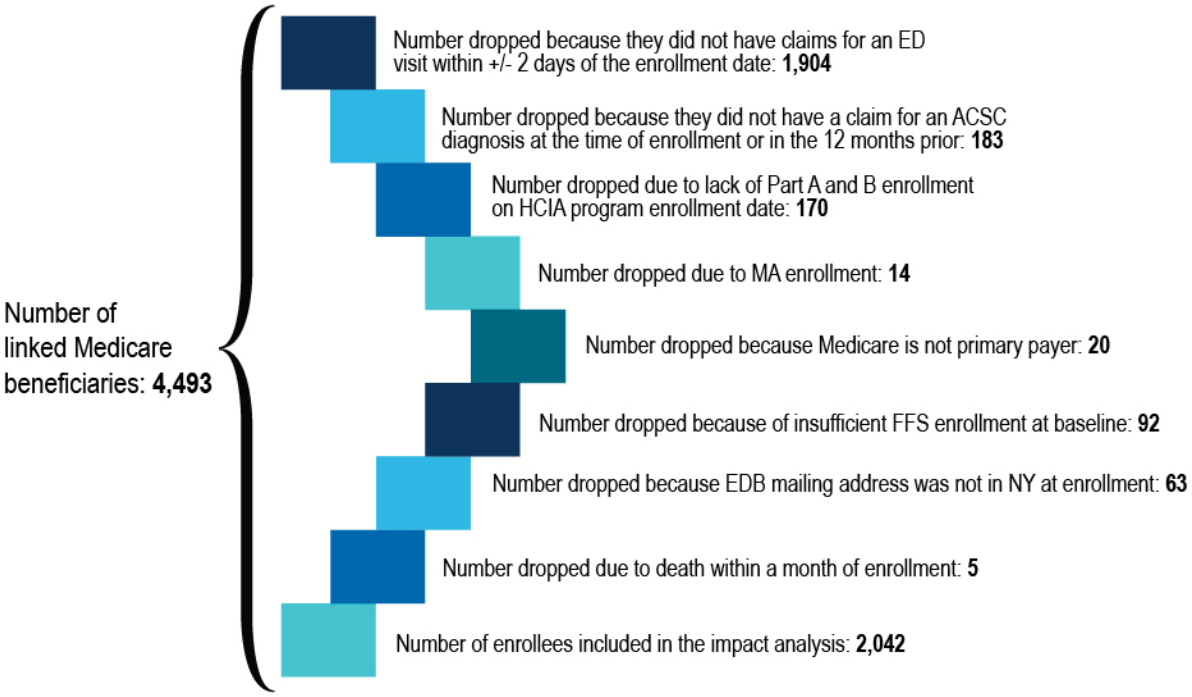
New York City Health + Hospitals started enrolling consenting participants in its ED Care Management program, including Medicare and Medicaid beneficiaries and those with other types of insurance, in September 2014. As of the end of August 2017, the program had 83,946 unique direct program participants.[3] The awardee estimates that about 29 percent of these participants are in Medicaid managed care, 9 percent are in Medicaid FFS, 7 percent are in Medicare managed care, and 6 percent are in Medicare FFS. Most of the remaining 49 percent are uninsured or privately insured, with a small proportion having other types of insurance such as workers' compensation.

In the baseline characteristics that are presented here, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date). They were also required to have met all of these program criteria for a period of at least 90 days during the baseline year (the 365 days immediately before their enrollment plus the date of their enrollment ED visit). In addition, they must have been enrolled in the awardee's program on or before November 30, 2016, in order to allow for at least six months of program exposure for interim impacts estimates and to ensure a long enough run-out period to capture nearly all claims for the most recent participants. Furthermore, in order to identify primary diagnosis at enrollment, we restricted the treatment group to beneficiaries for whom we could find emergency department claims at one of the intervention hospitals within two days before or after the beneficiary's enrollment date as reported by the awardee. Finally, in order to limit the potential for selection based on unobserved (in claims) characteristics to bias our analysis, we restricted the treatment group to beneficiaries who had a targeted ACSC as a primary or secondary diagnosis at the time of enrollment or during the 12 months before enrollment. The calendar period covered by the baseline quarters is based on the enrollment date for each participant, and will therefore vary by participant. After we excluded beneficiaries who

[3] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

did not meet the above criteria, a total of 2,042 participants were included in the analysis of baseline characteristics for this report.

Figure III.1 Reasons for exclusion of program enrollees from impact analyses



This chapter also presents (1) estimates of the prevalence of ACSCs; (2) the proportion of those without ACSCs in the baseline period; and (3) total Medicare expenditures per beneficiary per month (PBPM), stratified by each ACSC present among participants. We used a list of 16 targeted ACSCs and corresponding diagnosis codes that the awardee developed to identify and monitor the mix of participants in its program. Initially, the awardee targeted ACSCs based on the Prevention Quality Indicators (PQI) from the Agency for Healthcare Research and Quality. However, as the program progressed, the awardee modified and added to the PQIs in order to better identify the population that would benefit most from the intervention. We use the awardee's definitions here so our reporting can correspond to the way New York City Health + Hospitals identifies and classifies its program participants.

Our analysis of baseline characteristics indicates that the awardee is serving a demographically diverse population with significant health care needs (Table III.2) and high Medicare expenditures (Table III.3). Thirty-one percent of the program's participants are younger than 65, and 8 percent are 85 or older. Forty-one percent of the participants originally enrolled in Medicare because of a disability, which is significantly higher than the national rate of 24 percent (based on a 5 percent sample of Medicare beneficiaries from 2015). Two percent of participants have end-stage renal disease (ESRD).

Participants are more likely to be female (58 percent) than male. Only 30 percent of participants are white (significantly lower than the 80 percent of beneficiaries nationwide who are white)—reflecting the racial composition of the four New York City boroughs that the

participating hospitals are located in. Fifty-four percent of participants are dually eligible for Medicare and Medicaid, which suggests that they have a high level of social need; 18 percent of beneficiaries nationwide are dually eligible. Participants have substantially poorer health status and greater needs for care than the general Medicare FFS population, as evidenced by the fact that their average hierarchical condition categories (HCC) risk score is 52 percent higher than that of the average Medicare FFS beneficiary.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the treatment group of the New York City Health + Hospitals program through November 30, 2016

	All participants (n = 2,042)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	634	31
65 to 74	755	37
75 to 84	480	24
85 and older	173	8
Gender		
Female	1,191	58
Male	851	42
Race		
White	609	30
Black	875	43
American Indian, Alaska Native, Asian/Pacific Island American, or other	205	10
Hispanic	317	16
Original reason for Medicare eligibility		
Old age and survivor's insurance	1,167	57
Disability insurance benefits	830	41
End-stage renal disease (ESRD)[a]	45	2
Hospice[b]	1	0.05
Medicare/Medicaid dual status, percentage dual[b]	1,102	54
HCC score[c]		Statistic
Mean		1.52
25th percentile		0.73
Median		1.16
75th percentile		1.84

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the day after the beneficiary is discharged from the emergency department visit at which he or she enrolled. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

Table III.2 *(continued)*

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

The participants had high Medicare expenditures and high rates of service use in the year before they enrolled, which was consistent with their demonstrated needs. In Table III.3, we report baseline expenditure and utilization data for a common set of measures, including the four core measures specified by CMMI. We examined baseline cost of care by calculating average PBPM[4] Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $2,045—substantially higher than the 2014 national average of $792.[5] Average PBPM Medicare payments for inpatient ($1,148), outpatient and ED visits ($342), and physician visits ($288) were the largest drivers of total cost of care.

Participants had high average use of expensive Medicare services before the ED visit that precipitated enrollment in the awardee's program. The annual rate of acute care hospitalizations was 866 per 1,000 Medicare FFS beneficiaries in the treatment group during the baseline year—over three times the national annual average of 276 per 1,000 Medicare FFS beneficiaries in 2014. Furthermore, for program participants, the annual rate of ED visits not leading to a hospitalization was 3,628 per 1,000 Medicare FFS beneficiaries. Not surprisingly, the 2014 national annual rate of ED visits that did not lead to a hospitalization was considerably lower, at 454 per 1,000 Medicare FFS beneficiaries—a difference that reflects the program's recruitment of frequent ED users for care management. As required by our restrictions on the treatment group we analyzed, 100 percent of participants had an ambulatory ED visit in the last quarter of their baseline period.

The likelihood of a 30-day readmission for program participants (31 percent among discharges) was about 70 percent greater than the 2014 national average for Medicare FFS beneficiaries (18 percent). Thus, there is an opportunity to reduce potentially avoidable ED visits and readmissions in this population, both during the 90-day intervention period and beyond, through enhanced access to follow-up care. At baseline, the annual rate of primary care visits in any setting for program participants (5,015 per 1,000 Medicare FFS beneficiaries) was substantially lower than their rate of specialty service use in any setting (11,258 per 1,000 Medicare FFS beneficiaries). This may suggest a need for greater access to primary care, which is a focus of the program. On the other hand, given the health care needs of this population, the high rates of specialty care use may be appropriate.

[4] The months referred to in our calculations are 30-day periods rather than calendar months.

[5] The national data presented here and in the next paragraph are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

In the last baseline quarter we observed a steep increase in average PBPM total payments relative to the previous three quarters (32 percent) and also in average PBPM payments for inpatient (33 percent), outpatient (55 percent), and physician services (27 percent). We observed a similar trend in rates of hospitalizations, outpatient ED visits, primary care visits, and specialty services. The trend is not surprising, given that beneficiaries generally entered the program at the time of an ED visit. In sum, beneficiaries targeted for this intervention had high care expenses and high use of acute care both during the year before they enrolled and in the quarter immediately before they enrolled. This trend has important implications for selecting a comparable cohort of Medicare FFS beneficiaries as the comparison group.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the treatment group of the New York City Health + Hospitals program through November 30, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	2,042	1,887	1,943	2,035	2,042
Average Medicare expenditures PBPM[a]					
Total	2,045 (91)	1,875 (126)	1,860 (122)	1,905 (109)	2,489 (128)
Acute inpatient	1,148 (68)	1,041 (88)	1,041 (95)	1,074 (86)	1,402 (101)
Inpatient other[b]	86 (12)	60 (25)	90 (21)	94 (22)	97 (20)
Outpatient[c]	342 (16)	294 (24)	308 (21)	294 (18)	462 (20)
Physician services	288 (11)	272 (13)	265 (12)	269 (13)	340 (14)
Home health	58 (4)	53 (6)	57 (7)	52 (6)	68 (7)
Skilled nursing facility	109 (14)	138 (41)	88 (35)	112 (23)	100 (22)
Hospice	1 (1)	0 (0)	0 (0)	0 (0)	3 (3)
Durable medical equipment	13 (1)	17 (3)	11 (1)	10 (1)	17 (3)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	866 (54)	788 (65)	783 (63)	832 (62)	1,043 (69)
Outpatient ED visits[e]	3,628 (220)	2,261 (226)	2,087 (155)	2,535 (268)	7,309 (260)
Primary care visits in any setting	5,015 (191)	4,726 (239)	4,912 (234)	4,763 (240)	5,590 (246)
Primary care visits in ambulatory settings	3,671 (133)	3,533 (177)	3,594 (162)	3,573 (158)	3,943 (182)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Specialist visits in any setting	11,258 (415)	11,075 (479)	10,828 (469)	10,764 (464)	12,240 (504)
Specialist visits in ambulatory settings	7,422 (236)	7,499 (284)	7,534 (303)	7,163 (265)	7,472 (281)
Measures of any health care utilization					
Percentage with a hospital admission[d]	35 (1)	12 (1)	14 (1)	13 (1)	16 (1)
Percentage with an outpatient ED visit[e]	100 (<0.5)	24 (1)	24 (1)	25 (1)	100 (<0.5)
Percentage with a 30-day readmission among all discharges	31 (1)	30 (2)	26 (2)	30 (2)	33 (2)
Percentage of participants with a readmission among all participants	9 (1)	3 (<0.5)	3 (<0.5)	3 (<0.5)	4 (<0.5)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date plus the date of the enrollment ED visit. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby the four quarters sum to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient other expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

1. ACSCs among treatment group members and potential comparison group members

Because the initiative provides comprehensive care management to adult patients with ACSCs for 90 days after an ED visit, we examined the prevalence of the most common ACSCs among participating Medicare FFS beneficiaries. Table III.4 shows the prevalence of awardee-defined ACSCs that we identified in the baseline year among participating Medicare FFS beneficiaries and the corresponding average total Medicare expenditures PBPM. Due to restrictions we placed on the analytic treatment group, all participating Medicare FFS

beneficiaries in the treatment group had one or more ACSC, as identified through a search of inpatient, hospital outpatient, and carrier claims data from their baseline year. We also found that the three most common ACSCs among these Medicare FFS beneficiaries were (1) hypertension (77 percent), (2) diabetes (54 percent), and (3) chest pain (37 percent). The three most expensive categories of ACSCs (considering all expenditures for those with a given ACSC) for this group were (1) bacterial pneumonia ($6,371 average total Medicare expenditures PBPM); (2) sickle cell disease ($6,323 average total Medicare expenditures PBPM); and (3) heart failure ($5,561 average total Medicare expenditures PBPM). Only 1 percent of the participants had sickle cell disease, whereas 9 percent had bacterial pneumonia.

The average total cost PBPM for those flagged with an ACSC in the baseline year was almost nine times what it was for those without an ACSC. Furthermore, the costs for those with an ACSC, in all categories of expenditures that we examined, were higher than the costs arising from other participants' care. These differences are statistically significant at the 5 percent level in all categories of expenditures and utilization we examined, except expenditures in the other inpatient, hospice, and durable medical equipment categories. Similar to the patterns noted for the full group of participants, the patterns of increasing average expenditures and increasing utilization from the first three-quarters to the last baseline quarter held for those with any ACSC.

Table III.4. Prevalence of ACSCs and total average Medicare expenditures in the baseline year for Medicare FFS beneficiaries enrolled in the New York City Health + Hospitals program through November 30, 2016

ACSC type	Count of FFS beneficiaries	Percentage of total	Total average Medicare expenditures PBPM
Hypertension	1,575	77	$2,305
Diabetes	1,107	54	$2,309
Chest pain	761	37	$3,307
Asthma	509	25	$2,705
Urinary tract infection	467	23	$2,830
Intractable pain	303	15	$4,600
Chronic obstructive pulmonary disease or asthma in older adults	286	14	$3,989
Dehydration	288	14	$5,609
Heart failure	268	13	$5,561
Cellulitis	221	11	$4,490
Syncope	205	10	$3,762
Angina	178	9	$4,878
Bacterial pneumonia	176	9	$6,371
Seizure	167	8	$4,325
Deep vein thrombosis	62	3	$4,899
Sickle cell disease	30	1	$6,323
Any ACSC	2,042	100	$2,045

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Notes: Individual and composite ACSCs were identified by using the awardee's definitions applied to inpatient, hospital outpatient, and carrier claims data.

Table III.4 *(continued)*

Total Medicare expenditures were calculated from all claims for each participant with at least one eligible day during the baseline year or any baseline quarter. These measures are means for the individuals in the analysis sample, not institutional means.

ACSC = ambulatory care sensitive condition.

C. Quality of matched comparison group

In this section, we describe the matching process used to identify a comparison group.

1. Description and identification of the treatment group

In preliminary analyses of Medicare FFS participants in the New York City Health + Hospitals program, we found that 79 percent had an ACSC diagnosis either at the ED visit when they enrolled or during the 12 months before they enrolled. Because an ACSC diagnosis and an outpatient ED visit are the two enrollment criteria that can be observed in claims, we have chosen to restrict our analysis to these participants.[6] We divided the treatment and potential comparison groups into three subgroups based on whether each beneficiary had:

- An ACSC as a primary diagnosis at the time of the enrollment/pseudo-enrollment[7] ED/observation visit (Subgroup 1)
- No ACSC primary diagnosis, but an ACSC secondary diagnosis at the time of enrollment/pseudo-enrollment (Subgroup 2)
- No ACSC as a primary or secondary diagnosis at the time of enrollment/pseudo-enrollment, but an ACSC diagnosis (identified in any setting) during the 12 months prior (Subgroup 3).

The 16 ACSCs targeted by the New York City Health + Hospitals program are listed in Table III.5. In addition to targeting patients with these conditions, NCMs in the intervention enroll patients who meet utilization-based criteria and other patients they believe could benefit from the program.

[6] Of the remaining 21 percent of Medicare FFS participants, we found that nearly 74 percent (16 percent of the total) did not have a claim for an ED visit to an intervention hospital within two days of their enrollment date, while the other 26 percent (5 percent of the total) had a claim for an outpatient ED visit around the date of enrollment but not an ACSC diagnosis. In our last communications from New York City Health + Hospitals, the awardee expressed uncertainty about the precise cause of the missing ED or observation visit claims for participants. However, the awardee has been looking into it and believes it may be due to delays in submitting or processing claims, perhaps in situations where more clarification of information is needed to process the claim.

[7] For treatment group members, we define their enrollment date as the day after the ED or observation visit identified in claims that is closest to the enrollment date provided by the awardee. For potential comparison beneficiaries, we define a pseudo-enrollment date as the day after an ED or observation visit that did not result in an inpatient stay.

Table III.5. New York City Health + Hospitals ACSCs

ACSC name
Angina
Asthma
Cellulitis
Chest pain
Chronic obstructive pulmonary disease (COPD)
Deep vein thrombosis (DVT)
Dehydration
Diabetes
Heart failure
Hypertension
Intractable pain
Pneumonia
Seizure
Sickle cell
Syncope
Urinary tract infection (UTI)

Source: Communication from New York City Health + Hospitals.
ACSC = ambulatory care sensitive condition.

We separated beneficiaries into these three subgroups and estimated a propensity score model for each subgroup of beneficiaries. Within each subgroup, we exact-matched treatment and comparison beneficiaries based on the quarter of enrollment or pseudo-enrollment and an ACSC category from among 16 ACSCs targeted by the awardee.

2. Identifying a potential comparison group

The pool from which the propensity score–matched comparison group was drawn consisted of Medicare FFS beneficiaries who had outpatient ED or observation visits to any of nine comparison EDs in New York City and who met the ACSC-based inclusion criteria that define the three subgroups described above. We selected the nine comparison hospitals because they operate in the same five hospital service areas as the intervention hospitals, and they serve populations with characteristics (that is, percentage uninsured, poverty levels, and educational attainment) similar to the characteristics of people served by the intervention hospitals. Five of the nine comparison hospitals are other hospitals in the New York City Health + Hospitals system that are not participating in the HCIA R2–funded program. Also, we selected comparison hospitals that are similar to the intervention hospitals in terms of size (number of beds) and occupancy rate. Table III.6 lists the treatment and comparison hospitals and their characteristics.

Table III.6. Treatment and comparison hospitals, New York City Health + Hospitals

Hospital name	Hospital service area	Number of beds	Occupancy rate
Treatment hospitals			
Bellevue	Manhattan	476	0.66
Elmhurst	Flushing	350	0.78
Jacobi	Bronx	326	0.78
Kings	Brooklyn	406	0.83
Lincoln	Bronx	287	0.76
Queens	Jamaica	190	0.86
Comparison hospitals			
Coney Island	Brooklyn	261	0.90
Harlem	Manhattan	205	0.62
Metropolitan	Manhattan	181	0.59
North Central Bronx	Bronx	103	0.54
Woodhull	Brooklyn	238	0.61
New York Queens	Flushing	408	0.88
Flushing Medical Center	Flushing	275	0.84
Forest Hills	Flushing	216	0.88
Jamaica	Jamaica	316	0.79

Source: CMS 2014 Provider-of-Services File.

3. The matching process

Using the sample of beneficiaries who had outpatient ED visits to the hospitals in Table III. 6, we performed propensity score matching at the beneficiary level using the optimal matching algorithm separately in the three subgroups. The model specification was the same in all three subgroups and included as predictor variables the characteristics listed in the second column of Table III.7 below. We did not include quarter of enrollment/pseudo-enrollment and ACSC indicators (Column 1) in the propensity score model because we controlled for them through exact matching. For potential comparison facilities, we used a pseudo launch date of September 1, 2014, and assigned enrollment dates for potential comparison beneficiaries in the same fashion as for treatment beneficiaries. We defined and constructed pre-enrollment characteristics during the one-year period preceding the day after the enrollment/pseudo-enrollment visit occurred. Our sample consists of nine quarters with enrollment/pseudo-enrollment dates from September 1, 2014, through November 30, 2016.

We used calipers to further improve balance on certain beneficiary-level characteristics—(1) number of outpatient ED/observation visits in the previous year for all three subgroups, (2) primary diagnosis of lower respiratory conditions (an Agency for Healthcare Research and Quality [AHRQ] Clinical Classification Software [CCS] category) in Subgroup 2, and (3) primary diagnoses of schizophrenia and alcohol-related disorders in Subgroup 3.[8] That is, the

[8] We included indicators for AHRQ CCS categories of the primary diagnoses at time of enrollment for Subgroups 2 and 3 in order to account for the primary reason for the enrollment ED/observation visit among those beneficiaries

matching process imposed penalties for ED/observation visits greater than 2.6 standard deviations apart (based on treatment group standard deviation), and it imposed a penalty for non-matches on the AHRQ CCS categories for lower respiratory conditions in our second subgroup and penalties for certain psychiatric and alcohol-related disorders in the third subgroup. Thus, a potential comparison patient whose propensity score is the closest match to a particular treatment observation, but which differs from it on number of ED/observation visits by more than 7, may not be selected as one of the best matches for that treatment beneficiary.

Table III.7. New York City Health + Hospitals: variables used in exact matching and propensity score matching

Exact matching variables	Propensity score matching variables
All subsets	Age (continuous)
Quarters 1–9	Baseline HCC score[a]
Subset 1	Gender (Female)
16 primary ACSC diagnosis indicators	Race: white
Subset 2	Race: black
16 secondary ACSC diagnosis categories	Race: American Indian/Asian/Pacific Islander
Subset 3	Ethnicity: Hispanic
16 ACSC diagnosis indicators from the 12 months before enrollment	Hospital occupancy rate (hospital where enrollment visit took place)
	Medicare/Medicaid dual eligible
	Medicare expenditures during 12 months before enrollment
	Total Medicare expenditures[b] (continuous)
	Hospital outpatient expenditures (continuous)
	Physician services expenditures (continuous)
	Physicians per 100,000 residents (Hospital Service Area level)
	AHRQ Clinical Classification Software categories for primary diagnosis at the enrollment visit among beneficiaries without an ACSC as primary diagnosis during the enrollment visit
	Utilization during 12 months before enrollment
	Number of acute hospital admissions[c]
	Number of primary care visits - ambulatory
	Number of specialist (any type) visits - ambulatory
	Number of outpatient ED/observation visits
	Zip code uninsured proportion

[a]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS.

[b]Total Medicare expenditures for the baseline year were calculated from all claims for each participant with at least one eligible day during that year.

[c]The hospitalization measure includes acute care hospital admissions and excludes all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

ACSC = ambulatory care sensitive condition; AHRQ = Agency for Healthcare Research and Quality; HCC = hierarchical condition category.

in the treatment and comparison groups who were enrolled in the intervention or pseudo-enrolled in the comparison group due to a secondary or prior-year ACSC diagnosis. More information about the AHRQ CCS is available at https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp for use with ICD-9-CM diagnosis codes and at https://www.hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp for ICD-10-CM diagnosis codes.

We performed optimal matching with replacement, permitting matched sets comprising no more than 10 treated beneficiaries per matched comparison (10T:1C) and no more than 5 matched comparisons per treated beneficiary (1T:5C). We chose the lower bound (10T:1C) to limit the amount of replacement (and the corresponding high matching weights assigned to these comparison cases, which reduce precision); the upper bound (1T:5C) helps to exclude lower-quality matches from the sample.

For those with a targeted ACSC as a primary diagnosis at the time of the enrollment ED visit (Subgroup 1), this matching process yielded a total of 2,236 comparison beneficiaries matched with 828 treatment beneficiaries—an average of roughly three comparisons per treatment beneficiary. The distribution of treatment group beneficiaries by matching ratio (number of treatment beneficiaries matched to a comparison beneficiary) for Subgroup 1 is given in Table III.8.

For those with a targeted ACSC as a secondary diagnosis (and no ACSC for primary diagnosis) at the time of the enrollment ED/observation visit (Subgroup 2), this matching process yielded a total of 1,770 comparison beneficiaries matched with 502 treatment beneficiaries—an average of 3.4 comparisons per treatment beneficiary. The distribution of treatment group beneficiaries by matching ratio (number of treatment beneficiaries matched to a comparison beneficiary) for Subgroup 2 is given in Table III.8.

For those with a targeted ACSC diagnosed during the 12 months before enrollment and no ACSC for primary diagnosis or a secondary diagnosis at time of ED/observation visit (subgroup 3), the matching process yielded a total of 2,139 comparison beneficiaries matched with 712 treatment beneficiaries—an average of roughly 3 comparisons per treatment beneficiary. The distribution of treatment group beneficiaries by matching ratio (number of treatment beneficiaries matched to a comparison beneficiary) for Subgroup 3 is given in Table III.8. For the full sample combining the three subgroups, when taking into account the design effect due to weighting for the comparison group mean, the effective comparison group sample size is 2,615, which produces a ratio of 1.28 comparison members per treatment member.[9]

[9] The design effect is calculated as (number of comparison observations)*(sum of squared matching weights)/(square of the sum of matching weights). The effective comparison sample size is (number of comparison observations)/(design effect).

Table III.8. Matching ratios for three subgroups, New York City Health + Hospitals

Matching ratio (T:C)	Number of treatment beneficiaries		
	Subgroup 1: Matched on primary ACSC during ED visit	Subgroup 2: Matched on secondary ACSC during ED visit	Subgroup 3: Matched on ACSC, year before ED visit
7:1			7
6:1	6		
5:1	10		
4:1	28		8
3:1	30	12	18
2:1	66	18	60
1:1	239	115	190
1:2	56	26	46
1:3	43	22	36
1:4	47	21	25
1:5	303	288	322
	828	502	712

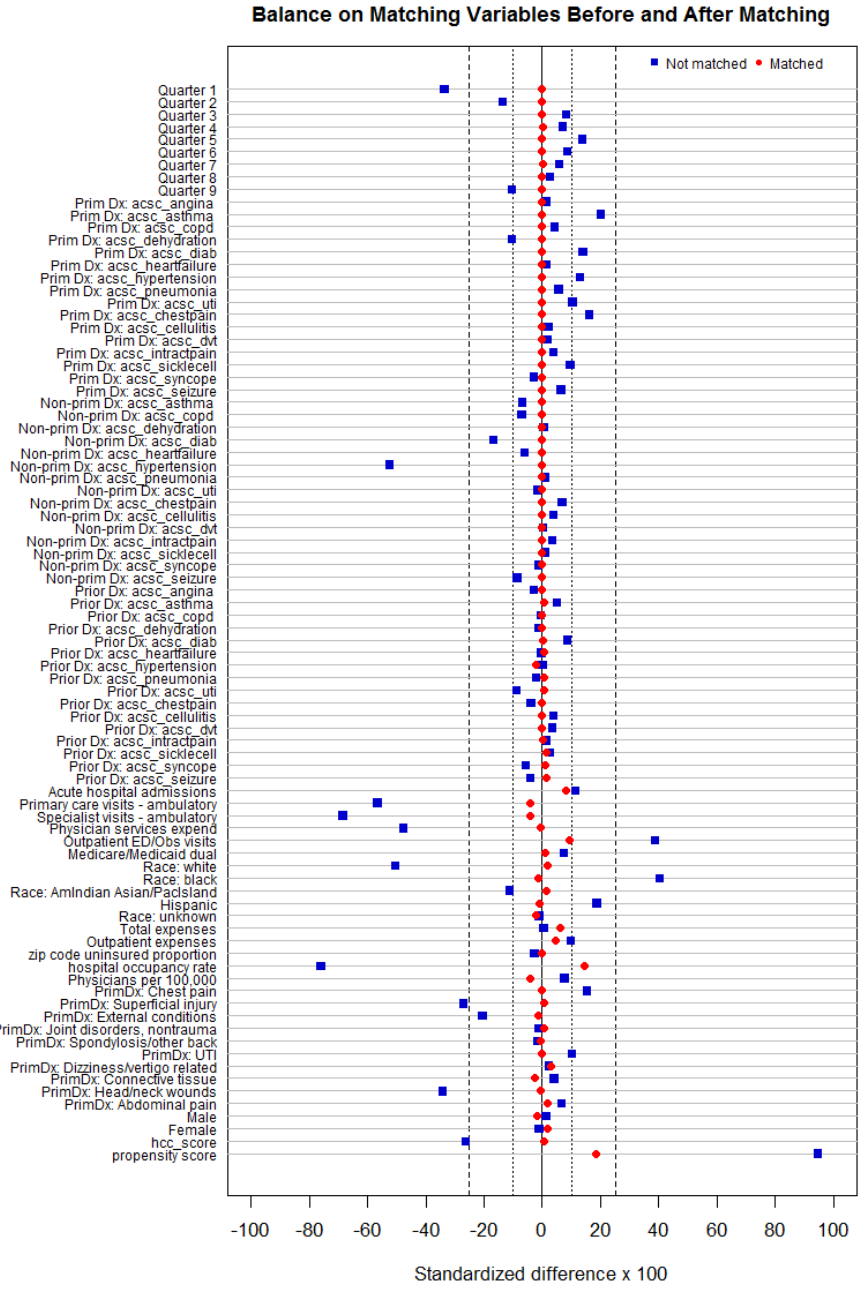
ED = emergency department; ACSC =ambulatory care sensitive condition.

For the full set of matched treatment and comparison beneficiaries, combining the three subgroups, we have 6,145 comparison beneficiaries matched with 2,042 treatment beneficiaries—an average of roughly three comparisons per treatment beneficiary. For the combined set of three subgroups, the standardized differences between the treatment and matched comparison groups for all matching covariates except one were less than 0.1 (see Figure III.2). The one variable that had a standardized difference between treatment and matched comparison groups greater than 0.1 was hospital occupancy rate, which had a standardized difference of 0.14. This still meets the standard threshold of 0.25 suggested by Rubin (2001), and the absolute difference in occupancy rates was relatively small.[10] The standardized difference in the mean propensity score values was 0.18, and the treatment and matched comparison groups had mean propensity score values of 0.23 and 0.20, respectively.

Our regression analysis to estimate program impacts controls for the remaining differences between the two groups on these characteristics. For the final report, we will reevaluate treatment-comparison group differences in mean outcomes at baseline with the final analytic study sample. We will also revisit the propensity score matching approach and reassess the need to impose additional calipers on key baseline outcome measures to ensure the baseline differences are minimized.

[10] Rubin, D. B. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology,* vol. 2, no. 3–4, 2001, pp. 169–188.

Figure III.2. Comparison of balance on matching variables before and after matching for the three subgroups combined, New York City Health + Hospitals



Note: The primary diagnoses shown near the bottom of the figure represent the 10 most common condition categories found in Subgroup 3, found by applying the AHRQ CCS to each beneficiary's primary diagnosis at the time of enrollment/pseudo-enrollment. Standardized differences between the treatment and matched comparison groups for all AHRQ CCS categories were less than 0.1. We show only the 10 most common CCS categories.

A full balance table and balance tables for each the three subgroups are in Appendix A of this report.

Prim Dx = primary diagnosis at enrollment/pseudo-enrollment visit; Non-prim Dx = non-primary diagnosis at enrollment/pseudo-enrollment visit; Prior Dx = diagnosis in any setting during the 12 months before the enrollment/pseudo-enrollment visit; acsc = ambulatory care sensitive condition; copd = chronic obstructive pulmonary disease; diab = diabetes; uti = urinary tract infection; dvt = deep vein thrombosis; intractpain = intractable pain; expend = expenditures.

D. Interim impact findings

As part of the program, an NCM creates an ambulatory care plan, provides health education, and links the participant to other services or providers as needed, such as primary care, a home care intake nurse, or a pharmacist. CLWs provide follow-up care coordination during the 90 days after a participant's qualifying ED visit. New York City Health + Hospitals hypothesized that giving patients with ACSCs who seek care in its EDs these comprehensive care management services and extended care coordination should reduce the incidence of preventable ED and hospital use.

1. Data and estimation approach

Our impact analyses include (1) Medicare FFS beneficiaries whom the awardee reported as participants with enrollment dates between September 1, 2014, and November 30, 2016, and (2) matched comparison beneficiaries with ED visits to comparable hospitals in New York City during the same time period. We estimated impacts on outcomes for three different post-enrollment follow-up periods: 6 months, 12 months, and 24 months. We did this to test the hypothesis that outcomes from services such as health education from the NCM and linkage to primary care providers would persist beyond the 90-day care management period for up to two years after enrollment in the program. For this interim analysis, we do not report the 6-month follow-up period results to allow for additional programmatic support if delays in coordinating care with primary care and specialty providers were encountered during the initial 90-day period. For this interim analysis, we restrict reporting to two analytic samples corresponding to the 12- and 24-month follow-up periods:

(1) 6,367 Medicare FFS beneficiaries (1,589 treatment and 4,778 comparison beneficiaries) who enrolled by May 31, 2016 to have 12 months of follow-up after enrollment

(2) 2,333 Medicare FFS beneficiaries (551 treatment and 1,782 comparison beneficiaries) who enrolled by May 31, 2015 to have 24 months of follow-up after enrollment.

Beneficiaries who died within a given follow-up period are included in the analyses for that period. We report findings for the 12-month follow-up period in the main part of the report; findings from the 24-month follow-up period are in Appendix B.

Our main data sources were the Medicare Enrollment Database (EDB) and Medicare FFS claims. We used Medicare enrollment data from September 1, 2013, through November 30, 2016, to identify Medicare program eligibility and beneficiaries' demographic characteristics. We used Medicare FFS claims from September 1, 2013, through May 31, 2017, to construct baseline and intervention period outcome measures and claims-based covariates. We allowed for three months of claims run-out and pulled the enrollment and claims data in September 2017.

Our framework was an intent-to-treat approach, and we followed intervention and comparison beneficiaries over time. We used a difference-in-differences model estimating the pre-post changes in outcomes for beneficiaries in the treatment group and for the matched comparison group over the same period, controlling for baseline beneficiary characteristics. The impact of the program is estimated as the difference between the average change over time for treatment beneficiaries and the average change over time for the matched comparison

beneficiaries. Thus, we attribute any changes observed in the difference between treatment and comparison beneficiaries' outcomes during the intervention period to the effects of the intervention.

The regression models control for all of the baseline beneficiary characteristics evaluated in Section III.C, and listed in Table III.7 with a number of exceptions: we excluded outcome variables (that is, expenditures and utilization) from the models, and we included indicator variables for only 24 AHRQ CCS categories with 100 or more cases. In addition, the regression models included the following variables: the baseline proportion of the population from the beneficiary's zip code living in poverty, and the baseline proportion of the adult population from the beneficiary's zip code with a high school degree. We also included indicator variables for each of the 6 treatment hospitals and 9 comparison hospitals in the models. The impact methodology (included as Appendix B to the main body of this report) describes the modeling approach in detail.

2. Outcomes

We examined the effects of the New York City Health + Hospitals program on a set of outcomes listed in Table III.9. We examined changes between treatment and comparison groups during the first 12 and 24 months of program follow-up in (1) PBPM Medicare FFS expenditures, in total, for acute inpatient care; other inpatient care; outpatient services (including ED visits); physician services; home health services; skilled nursing facility services; hospice services; and durable medical equipment (DME); (2) rates of hospitalization, ED visits/observation stays, primary care visits in ambulatory settings, specialist visits in any setting, and specialist visits in ambulatory settings per 1,000 beneficiaries; and (3) likelihood of a hospital admission, ED visit/observation stay, and a 30-day unplanned readmission among all beneficiaries.

Table III.9. Outcome measures, New York City Health + Hospitals

Outcome	
Average Medicare expenditures PBPM	
Total	Home health
Acute inpatient	Skilled nursing facility
Inpatient other	Hospice
Outpatient	Durable medical equipment
Physician services	
Health care utilization rates (annualized per 1,000)	
Acute hospital admissions	Primary care visits in ambulatory settings
Outpatient ED visits and observation stays	Specialist visits in any setting
	Specialist visits in ambulatory settings
Measures of any health care utilization	
Percentage with a hospital admission	Percentage of beneficiaries with a 30-day readmission among all beneficiaries
Percentage with an outpatient ED visit or observation stay	

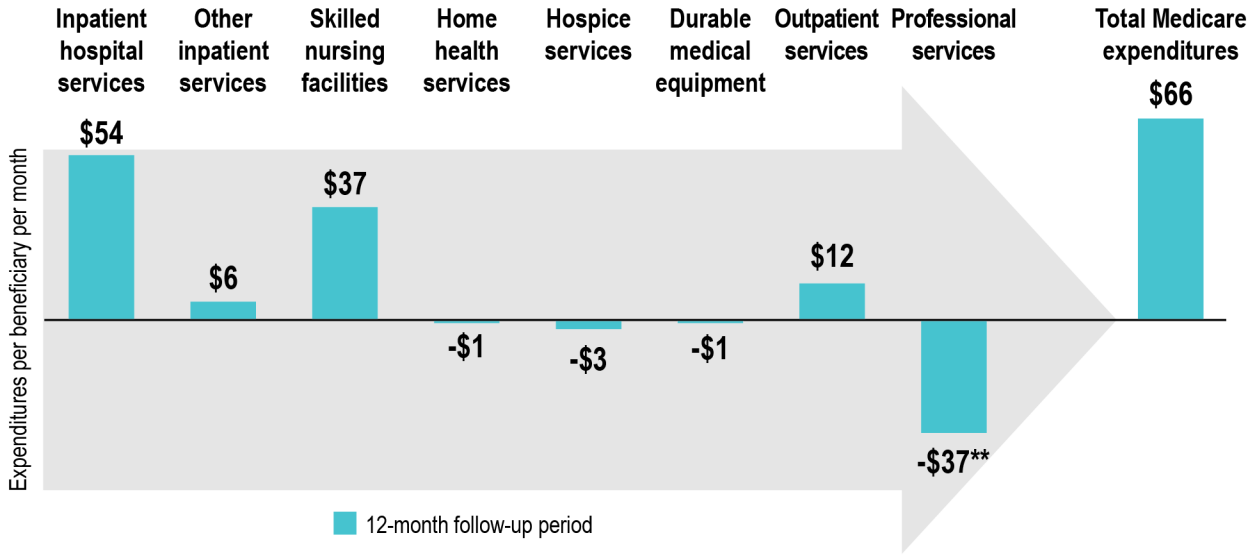
ED = emergency department.

3. Impact estimates

a. Medicare FFS expenditures

Only one of our estimates of program impacts on expenditure categories was statistically significant-- payments for physician services declined by $37 PBPM (90 percent confidence interval, or CI, = [-$58, -$16]), or 10 percent of the $376 that we estimate would have been observed for participants in the absence of the intervention (Figure III.3). Reducing use of physician services was not an intended outcome of the intervention. Total Medicare expenditures and spending for acute inpatient, other inpatient, outpatient, and skilled nursing facility services increased by 3 to 22 percent more for the treatment group than the matched comparison group, but none of the differences are statistically significant at the 10 percent confidence level. Expenditures for DME, home health, and hospice were lower by 1 to 17 percent, but none of these differences are statistically significant at the 10 percent confidence level.

Figure III.3. Estimated impact on total PBPM expenditures and by expenditure category, New York City Health + Hospitals



Source: Mathematica analysis of information from awardee's finder file through May 2016 and Medicare claims and enrollment data through September 30, 2017.

Notes: Number of treatment beneficiaries = 1,589; number of comparison beneficiaries = 4,778.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

PBPM = per member per month.

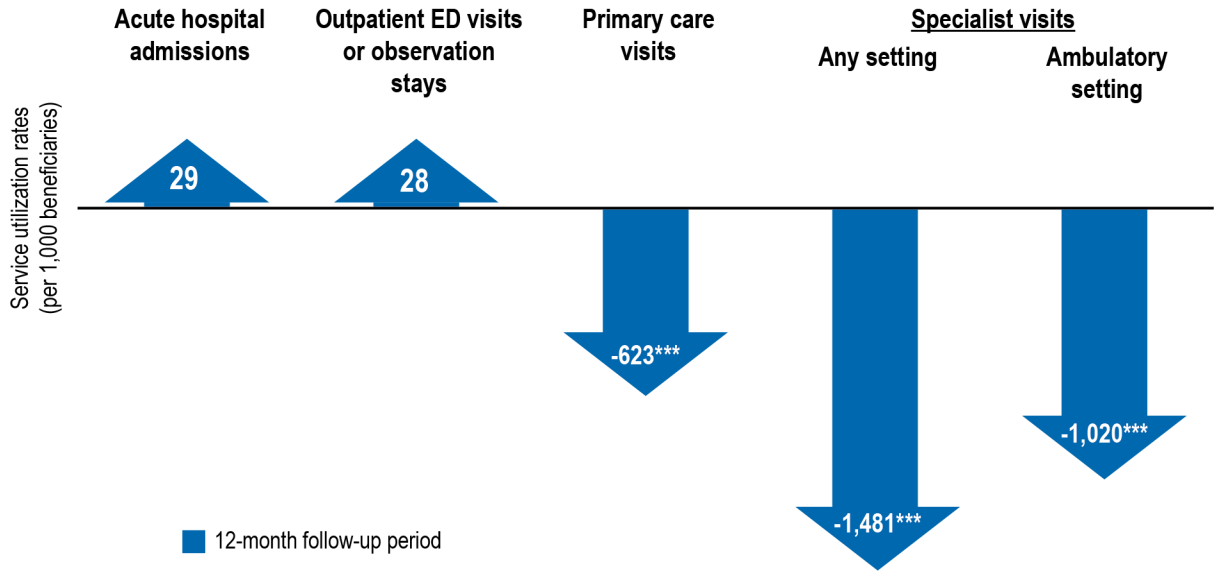
b. Service utilization

None of our impact estimates for the core utilization outcomes of hospitalizations, ED visits, and 30-day unplanned readmission were statistically significant (Figures III.4 and III.5); however, primary care visits and specialist visits (in any setting and in ambulatory settings) were significantly lower for the treatment group. The rate of primary care visits per 1,000 beneficiaries was -645 (90 percent CI = [-1,086, -204]), or 13 percent, lower than enrollees' predicted mean

(4874) in the absence of the intervention. The rates of specialist visits in any setting per 1,000 beneficiaries was lower by -1,481 visits (90 percent CI = [-2,358, -605]), which represents a 10 percent decline, and the rate of specialist visits in ambulatory settings per 1,000 beneficiaries was lower by -1,020 visits (90 percent CI = [-1,616, -424]), which represents an 11 percent decline.

On one outcome, we found evidence that the New York City Health + Hospitals program had the opposite of its intended impact. Our analysis of the program's impact on the probability of any ED or observation visit revealed that the program led to an increase of 2.2 (90 percent CI = [0.2, 4.3]) percentage points above the 42 percent visit rate that would be predicted absent the intervention.

Figure III.4. Estimated impact on rate of service utilization per 1,000, New York City Health + Hospitals



Source: Mathematica analysis of information from awardee's finder file through May 2016 and Medicare claims and enrollment data through September 30, 2017.

Notes: Number of treatment beneficiaries = 1,589; number of comparison beneficiaries = 4,778.

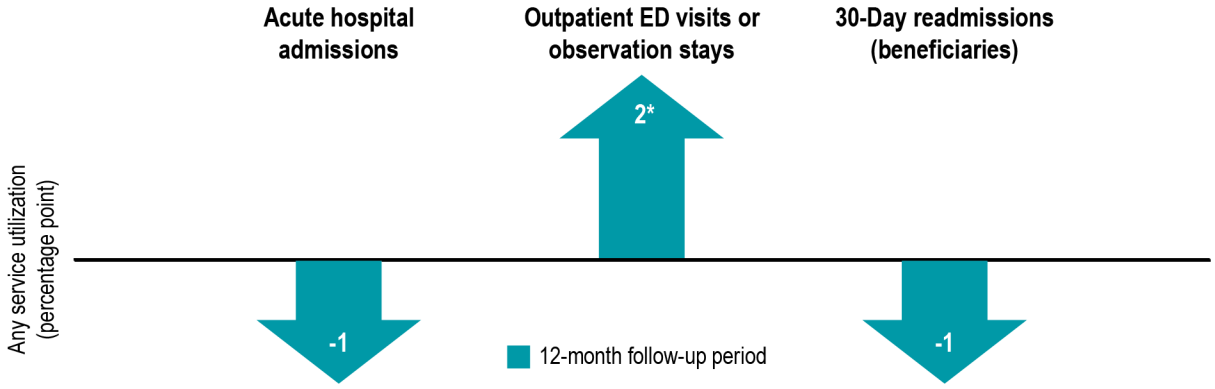
*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department.

Figure III.5. Estimated impact on likelihood of any service utilization, New York City Health + Hospitals



Source: Mathematica analysis of information from awardee's finder file through May 2016 and Medicare claims and enrollment data through September 30, 2017.

Notes: Number of treatment beneficiaries = 1,589; number of comparison beneficiaries = 4,778.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department.

4. Sensitivity analyses

We verified the robustness of our impact estimates by conducting two sets of sensitivity analyses: (1) extending the baseline period from one to two full years; and (2) topcoding outcomes to account for outliers with extreme values and re-estimating the regression models. Neither of the sensitivity analyses produced results that differed substantively from those presented in this report. In future analyses, we will conduct subgroup analyses if sample sizes are sufficiently large to produce statistically meaningful results.

5. Discussion: New York City Health + Hospitals

Our interim impact findings suggest that for the full sample of Medicare FFS treatment beneficiaries analyzed, the New York City Health + Hospitals program did not reduce use of acute care services and the emergency department as it was intended to. However, our estimates suggest that the program reduced ambulatory care primary care physician visits and specialist care visits, contrary to expectations. This reduction in ambulatory care reflects the challenges described in the implementation evaluation findings (Section II.B) that the program had with establishing linkages between participants and ambulatory care providers. Alternatively, this decrease in ambulatory care use, although it may be contrary to the program's intended goals of strengthening contact between beneficiaries and ambulatory health care services, may also suggest that other services the program provides, such as health education, obviate the need for ambulatory care from physicians. Also, a potential confounding issue related to primary care linkage arises from the fact that the five New York City Health + Hospitals sites in our comparison group participate in an ED care triage program funded through the DSRIP program. However, New York City Health + Hospitals did not implement the care management

component of the DSRIP-funded program until after the conclusion of the HCIA R2–funded program, and only a primary care linkage component of the DSRIP-funded program began in March 2016, more than six quarters after the HCIA R2–funded program began. Finally, this program required beneficiary consent which may lead to some degree of selection bias in our estimates. Further analyses will be conducted to assess the degree to which selection bias may be present, and if so, addressed quantitatively.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Instead of developing a stand-alone payment model for its program services, New York City Health + Hospitals is incorporating its program into its existing contracting vehicles. These vehicles include global risk capitated contracts with Medicaid and Medicare managed care plans and a value-based payment approach that aligns with broader New York Medicaid reform efforts. The awardee is also in the early stages of exploring shared savings through an all-payer ACO.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

New York City Health + Hospitals has identified three payment model options for continuing program operations. The first option builds on the awardee's existing global risk capitated contracts with two of its largest payers, MetroPlus (a Medicaid managed care organization [MCO]) and Healthfirst (a Medicaid MCO and Medicare Advantage plan). The program has been incorporated into these existing contracting vehicles. The second option, which would include other Medicaid managed care and commercial contracts, involves a value-based payment in which FFS payments would be adjusted for quality performance. For this option, the awardee is attempting to align with DSRIP, which seeks to reduce avoidable hospital use and requires the state to have 80 percent of its payments between Medicaid MCOs and providers in value-based payment arrangements by April 1, 2018. The third option—which is in the preliminary development stages—of developing an all-payer ACO leverages the awardee's current experience as an ACO in the Medicare Shared Savings Program and would allow the awardee to enter into a shared savings approach with the state Medicaid program.

C. Status of the payment model

New York City Health + Hospitals has executed its plans to incorporate its program into its existing global capitated arrangements with Healthfirst and MetroPlus. This financial arrangement is continuing after the end of the cooperative agreement. Both payers have provided claims data that have allowed the awardee to conduct a comprehensive cost-savings analysis of the financial impact of the care management strategy being offered through the program. The analyses based on MetroPlus data revealed a reduction in per-member per-month costs and utilization for program participants versus a control group of propensity-matched plan enrollees who were eligible for the program but were not enrolled. The awardee is conducting similar analyses based on the Healthfirst data. The awardee expects these analyses to give payers insight into the effects of the care management strategies on utilization and costs among the payers' members.

At the time of our interviews with the awardee, value-based FFS payment models were in preliminary negotiations, and the all-payer shared savings model was in an exploratory stage. New York City Health + Hospitals currently is in negotiations with commercial and Medicaid managed care payers with the goal of working with payers to create financial incentives to reduce ED and inpatient use through provision of care management and outpatient care services. Efforts to develop an all-payer ACO have been a lower priority and were in the exploratory stage during our interviews. Both of these efforts would apply to all services provided by New York City Health + Hospitals; they are not exclusive to the program services provided under the cooperative agreement.

D. Factors associated with the development of the payment model

Because the awardee primarily has been working through existing contractual arrangements, its focus has been less on developing a payment model and more on evaluating its current arrangements with health plans. The cooperative agreement has enabled the awardee to conduct analyses using claims data from payers that otherwise are not available to the awardee, which did not have access to the claims for the services it did not provide. For instance, the claims provided by Healthfirst includes medical, behavioral health, substance use, and pharmacy claims, regardless of which provider submitted the claims. The awardee also reported benefiting from learning activities sponsored by the implementation and monitoring contractor with other awardees, which promoted "outside-the-box" thinking and were an opportunity to hear how others have dealt with data challenges. The biggest barrier to the awardee's efforts has been harnessing its own clinical and financial data from its EMR and claims. There has been a learning curve in terms of understanding the limitations and caveats of the data and the reasons behind the data lags.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

New York City Health + Hospitals plans to sustain and scale its program after the cooperative agreement period by implementing a similar DSRIP-funded program that builds on the HCIA R2–funded program. The sustained program will be different in some ways, with changes based on recommendations and feedback from internal stakeholders. The awardee implemented the HCIA R2–funded program at six of its hospitals, and plans to scale the program to all 11 hospitals after the cooperative agreement ends. The awardee did not report plans to replicate its program.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, New York City Health + Hospitals planned to use lessons learned from the HCIA R2–funded program to inform a similar program funded by DSRIP. The awardee also planned to leverage global risk contracts with two large managed care organizations to cover program services; these contracts predated the award. The awardee also reported it was considering expanding the payment model to include Medicaid FFS, commercial payers, and uninsured patients.

C. Implementing the SSR plan: progress and changes

Sustainability. In the third program year, New York City Health + Hospitals planned to sustain the program after the cooperative agreement period ended by implementing a DSRIP-funded program that builds on it and is called ED Care Triage. New York City Health + Hospitals is also implementing other delivery system transformation projects funded by New York State's Medicaid DSRIP program.

During the third program year, the awardee partnered with a group of graduate students to develop a sustainability plan by researching care management models and best practices, interviewing OneCity leaders (OneCity Health is the New York City Health + Hospitals-sponsored Performing Provider System formed under the auspices of New York State's DSRIP program), and developing tools to assess staff competency, social determinants, and patient health literacy. The awardee informed its sustainability planning using proposals by the graduate students on how to fold the HCIA R2–funded program into the ED Care Triage program funded by DSRIP.

In the beginning of the third program year, the executive director of OneCity Health Care Management joined the team as the in-kind deputy executive sponsor. Previously, the award was overseen by the Division of Medical and Professional Affairs under the central office of New York City Health + Hospitals. According to a program leader, this transition enabled the awardee to better leverage lessons learned and inform sustainability planning. The executive sponsor and other program leaders used workgroup meetings to discuss successes, challenges, best practices, and trainings of interest to inform sustainability planning. Pharmacists, physician advisors, and operational staff (which included NCM leaders) each had separate workgroup meetings. In addition, physician advisors and leaders discussed eligibility criteria and how to integrate the DSRIP-funded program with other hospital initiatives.

Site visits by the project manager in the third year also informed sustainability planning. These site visits enabled leaders to develop a better understanding of similarities across facilities as well as the challenges unique to sites. Leaders leveraged information they learned on these site visits for DSRIP planning conversations, particularly conversations about eligibility criteria and adding a social worker, because they received a lot of feedback on the need for this role.

Finally, during the third program year, OneCity worked with New York City Health + Hospital's human resources and labor relations departments, and the Office of Transformation, to prepare transition plans for program staff.

Changes to the sustained program model. The new program model is similar to the program implemented during the cooperative agreement, but there are some differences. The positions of NCMs, CLWs, and pharmacists will be sustained after the cooperative agreement. Under the DSRIP-funded program, there will be multiple teams per site with expanded hours, instead of one team per ED. NCMs will play a more active role in communicating alternatives to admission to the ED providers. In addition, based on stakeholder feedback, the awardee decided on a 14-day follow-up period with participants, but program leaders are open to adjusting it once the program is underway.

The awardee also decided to adapt the program to better address the social determinants of health. First, the role of CLW will be expanded to include time to meet with higher-risk patients in the community, including home visits and escorting patients to appointments. Second, the new program model will include full-time licensed clinical social workers on the care teams. Third, the new program model will likely have expanded eligibility criteria to allow more focus on patients with behavioral health conditions, substance abuse, and social issues such as homelessness. These changes were based on recommendations by program staff who emphasized the need for the program to better address social determinants of the target population's health. The awardee's analytic work (identifying characteristics of highest-risk patients and the geographic concentrations of these patients) also informed sustainability planning about the social determinants of health.

The care teams under the DSRIP-funded program will also no longer include formal roles for the physician advisor or home care intake nurse. However, one program leader noted that physicians will informally continue championing and advising the program. Similarly, home care intake nurses will still be present in the EDs and serve as a resource for teams.

Scalability. New York City Health + Hospitals also reported plans to implement the DSRIP-funded program at all 11 of its hospitals after the cooperative agreement ends. During the cooperative agreement, only six of the hospitals implemented the program. Program leaders reported that the DSRIP-funded program would still have standardized policies, procedures, and metrics, but give local staff the flexibility to make adjustments based on the needs of their site and target population.

For example, the awardee reported toward the end of Year 3 that it would ask physician advisors and site leads to reassess the target population defined by the HCIA R2–funded project and adjust the inclusion criteria to better serve each hospital's patient population. In addition, all sites implementing the DSRIP-funded program will use the interoperable GSI platform to communicate and record information about care management. The awardee believes that information about care management will be better integrated across settings as a result. The awardee also planned to train staff for the DSRIP-funded program to supplement their knowledge and reinforce techniques and processes used by care management staff during the HCIA R2 project.

Replicability. In the third program year, New York City Health + Hospitals did not report plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

Program leaders and staff believed the features of the new program model would facilitate its sustainment. For example, a number of respondents noted the benefits of having CLWs meet participants in the community. This could alleviate the challenges of trying to reach participants by phone and also address the social determinants of health.

> "If you don't have buy-in from physicians, it will not be easy for care managers to push things forward."
>
> —Program staff

In addition, program leaders believed that scaling the adapted program model to all 11 sites would create a cultural shift that will facilitate sustainment at all sites. Program leaders were confident that implementing the program across the entire organization would increase its prominence, making clinicians and other staff more likely to support it. A number of the program staff we interviewed worried that eliminating the physician advisor role would keep clinicians and staff from engaging with the program. According to program staff, physician advisors were instrumental in increasing program awareness and buy-in among ED clinicians, providing medical expertise, and facilitating linkages to ambulatory care and home care. However, with the expected cultural shift across facilities, program leaders did not think that physician advisors required formal roles on the teams. Instead, leaders expected that physician advisors would continue to champion the program and provide advisory support, albeit informally, under the DSRIP-funded program.

Program leaders and staff expected primary care capacity issues to continue to challenge the program's ability to secure quick primary care appointments for patients. Physician advisors interviewed also highlighted access issues at community-based organizations (such as those serving individuals with disabilities), noting that such facilities often do not offer evening hours. Because these facilities do not fully meet the needs of the community, staff explained that patients will probably keep coming back to the ED until these other issues have been resolved.

Despite these concerns about access to primary care and community-based services, leaders at New York City Health + Hospitals are invested in strengthening their care management programs and the linkages across settings. Program leaders highlighted the strong support they have received from senior leaders at New York City Health + Hospitals, who view care management as the future of health care. In meetings with program staff and in our interviews, the awardee has described its vision of developing an integrated care management model across inpatient, ED, and ambulatory care settings. Further, one program leader noted that New York City Health + Hospitals, under a separate initiative, was in the process of launching call centers for every borough of the city, and hoped this would improve connectivity to ambulatory care in the future.

Finally, program leaders noted that staffing the new program and finalizing plans to transition staff from the HCIA R2–funded program to the DSRIP-funded program had been challenging, but they were gaining momentum at the time of our interviews. They explained that hiring processes in the hospital system were complex, and the transition required them to work closely with human resources and labor relations departments and with unions. Conversations were often delayed or required multiple rounds of meetings to move forward.

Program leaders acknowledged that the slow process stalled their ability to start the DSRIP-funded program, as the teams from the HCIA R2–funded program would be the champions of the new model and staffing needed to be finalized before the program could be implemented. Program leaders also noted their concern that the delayed timeline would affect training schedules. They further worried that delays in and uncertainty about the transition impeded their ability to clearly communicate with staff about transition plans, which could in turn dampen staff morale and might cause attrition.

However, despite the delay in communications and some signs of waning staff morale (revealed in interviews), one program leader noted recent progress in engaging stakeholders and taking concrete steps to implement a comprehensive communication plan, which included involving senior leaders in the organization to communicate with the sites. Although the communication plan to inform stakeholders, hospital leaders, and staff was moving forward, program leaders acknowledged that social workers still needed to be hired and that the complex hiring processes (which had also been a strong barrier during HCIA R2 implementation) remained a formidable challenge.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

The conclusions from the impact findings should be regarded as preliminary because the treatment group consists only of beneficiaries who entered the New York City Health + Hospitals program through November 2016. In the next analysis, we will include all beneficiaries who enrolled through May 2017 to allow for 90 days of program exposure before the program ended on August 31, 2017. We will also restrict enrollment to those before February 2017 to also test a longer program exposure window, 6 months as was tested in this interim analysis. With the final study sample, we will re-draw our matched comparison group and re-estimate the impact regression models to arrive at final impact estimates. The larger sample can be expected to result in smaller standard errors and therefore provide more precise estimates of program effects. The model will be estimated for a longer period of program operations, permitting us to estimate the effects, if any, of program maturity, and for subgroups of beneficiaries based upon presence of specific ACSCs. Special attention in the rematching may be necessary for these key subgroups of participants.

We will also perform more extensive sensitivity and robustness checks on all results, and expand our analyses of outcomes to include changes in the rate of ED visits and hospitalizations with a primary or secondary diagnosis of one of the 16 targeted ACSCs, and for ED visits within 90 days of program enrollment. In addition, we will estimate program impacts in a Bayesian framework, permitting the estimation of the program's impact to borrow strength from estimates from other time periods and outcomes, and also allowing us to estimate the probability that a program impact exceeds a specified statistical threshold.

We are also working toward incorporating into our evaluation Medicaid participants in the ED Care Management program. New York City Health + Hospitals has enough participation by Medicaid beneficiaries to detect impacts as small as 10 percent on some core outcome measures. We will continue to monitor the availability of Alpha-MAX and T-MSIS data for the entire intervention period with the goal of evaluating impacts on Medicaid. We will provide CMS with final impact results for New York City Health + Hospitals' Medicare and Medicaid participants as they become available in the upcoming year.

This page has been left blank for double-sided copying.

APPENDIX A

POST-MATCHING DIAGNOSTICS

This page has been left blank for double-sided copying.

Table A.1. Matching covariate balance: Full sample, New York City Health + Hospitals

Measure	Potential comparisons (n = 26740)	Matched comparisons (n = 6145)	Treated (n = 2042)	Adjusted difference (matched)	Percentage difference (matched)	Standardized difference (matched)	*p*-value (matched)	Equivalence *p*-value (matched)
Quarter 1 (proportion)	0.175 (0.002)	0.083 (0.004)	0.083 (0.006)	0.000 (0.008)	0.000	0.000	1.000	0.000
Quarter 2 (proportion)	0.143 (0.002)	0.101 (0.004)	0.101 (0.007)	0.000 (0.009)	0.000	0.000	1.000	0.000
Quarter 3 (proportion)	0.131 (0.002)	0.162 (0.005)	0.161 (0.008)	-0.001 (0.012)	-0.608	-0.003	0.933	0.000
Quarter 4 (proportion)	0.121 (0.002)	0.145 (0.004)	0.146 (0.008)	0.001 (0.011)	0.671	0.003	0.932	0.000
Quarter 5 (proportion)	0.105 (0.002)	0.155 (0.005)	0.155 (0.008)	0.000 (0.012)	0.000	0.000	1.000	0.000
Quarter 6 (proportion)	0.089 (0.002)	0.116 (0.004)	0.116 (0.007)	0.000 (0.010)	-0.424	-0.002	0.960	0.000
Quarter 7 (proportion)	0.086 (0.002)	0.103 (0.004)	0.103 (0.007)	0.000 (0.010)	0.474	0.002	0.960	0.000
Quarter 8 (proportion)	0.080 (0.002)	0.087 (0.004)	0.087 (0.006)	0.000 (0.009)	0.000	0.000	1.000	0.000
Quarter 9 (proportion)	0.070 (0.002)	0.048 (0.003)	0.048 (0.005)	0.000 (0.007)	0.000	0.000	1.000	0.000
Prim Dx: acsc_angina (proportion)	0.001 (0.000)	0.002 (0.000)	0.002 (0.001)	0.000 (0.001)	0.000	0.000	1.000	0.000
Prim Dx: acsc_asthma (proportion)	0.013 (0.001)	0.061 (0.003)	0.061 (0.005)	0.000 (0.007)	0.000	0.000	1.000	0.000
Prim Dx: acsc_copd (proportion)	0.008 (0.001)	0.012 (0.001)	0.012 (0.002)	0.000 (0.004)	0.000	0.000	1.000	0.000
Prim Dx: acsc_dehydration (proportion)	0.012 (0.001)	0.005 (0.001)	0.005 (0.002)	0.000 (0.002)	0.000	0.000	1.000	0.000
Prim Dx: acsc_diab (proportion)	0.015 (0.001)	0.043 (0.002)	0.043 (0.004)	0.000 (0.006)	0.000	0.000	1.000	0.000
Prim Dx: acsc_heartfailure (proportion)	0.003 (0.000)	0.003 (0.001)	0.003 (0.001)	0.000 (0.002)	0.000	0.000	1.000	0.000
Prim Dx: acsc_hypertension (proportion)	0.023 (0.001)	0.052 (0.003)	0.052 (0.005)	0.000 (0.007)	0.000	0.000	1.000	0.000
Prim Dx: acsc_pneumonia (proportion)	0.006 (0.000)	0.012 (0.001)	0.012 (0.002)	0.000 (0.003)	0.000	0.000	1.000	0.000
Prim Dx: acsc_uti (proportion)	0.029 (0.001)	0.052 (0.003)	0.052 (0.005)	0.000 (0.007)	0.000	0.000	1.000	0.000

Table A.1 *(continued)*

Measure	Potential comparisons (n = 26740)	Matched comparisons (n = 6145)	Treated (n = 2042)	Adjusted difference (matched)	Percentage difference (matched)	Standardized difference (matched)	*p*-value (matched)	Equivalence *p*-value (matched)
Prim Dx: acsc_chestpain (proportion)	0.044 (0.001)	0.091 (0.004)	0.091 (0.006)	0.000 (0.009)	0.000	0.000	1.000	0.000
Prim Dx: acsc_cellulitis (proportion)	0.016 (0.001)	0.019 (0.002)	0.019 (0.003)	0.000 (0.004)	0.000	0.000	1.000	0.000
Prim Dx: acsc_dvt (proportion)	0.001 (0.000)	0.002 (0.000)	0.002 (0.001)	0.000 (0.001)	0.000	0.000	1.000	0.000
Prim Dx: acsc_intractpain (proportion)	0.003 (0.000)	0.006 (0.001)	0.006 (0.002)	0.000 (0.002)	0.000	0.000	1.000	0.000
Prim Dx: acsc_sicklecell (proportion)	0.001 (0.000)	0.010 (0.000)	0.010 (0.002)	0.000 (0.003)	0.000	0.000	1.000	0.000
Prim Dx: acsc_syncope (proportion)	0.024 (0.001)	0.020 (0.002)	0.020 (0.003)	0.000 (0.004)	0.000	0.000	1.000	0.000
Prim Dx: acsc_seizure (proportion)	0.008 (0.001)	0.016 (0.001)	0.016 (0.003)	0.000 (0.004)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_asthma (proportion)	0.016 (0.001)	0.009 (0.001)	0.009 (0.002)	0.000 (0.003)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_copd (proportion)	0.008 (0.001)	0.004 (0.001)	0.004 (0.001)	0.000 (0.002)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_dehydration (proportion)	0.004 (0.000)	0.004 (0.001)	0.004 (0.001)	0.000 (0.002)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_diab (proportion)	0.102 (0.002)	0.062 (0.003)	0.062 (0.005)	0.000 (0.008)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_heartfailure (proportion)	0.004 (0.000)	0.001 (0.000)	0.001 (0.001)	0.000 (0.001)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_hypertension (proportion)	0.278 (0.003)	0.112 (0.005)	0.112 (0.007)	0.000 (0.010)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_pneumonia (proportion)	0.001 (0.000)	0.001 (0.000)	0.001 (0.001)	0.000 (0.001)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_uti (proportion)	0.012 (0.001)	0.010 (0.001)	0.010 (0.002)	0.000 (0.003)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_chestpain (proportion)	0.011 (0.001)	0.021 (0.002)	0.021 (0.003)	0.000 (0.004)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_cellulitis (proportion)	0.003 (0.000)	0.006 (0.001)	0.006 (0.002)	0.000 (0.002)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_dvt (proportion)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.001)	0.000	0.000	1.000	0.000

Table A.1 *(continued)*

Measure	Potential comparisons (n = 26740)	Matched comparisons (n = 6145)	Treated (n = 2042)	Adjusted difference (matched)	Percentage difference (matched)	Standardized difference (matched)	*p*-value (matched)	Equivalence *p*-value (matched)
Non-prim Dx: acsc_intractpain (proportion)	0.004 (0.000)	0.007 (0.001)	0.007 (0.002)	0.000 (0.003)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_sicklecell (proportion)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.001)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_syncope (proportion)	0.005 (0.000)	0.004 (0.001)	0.004 (0.001)	0.000 (0.002)	0.000	0.000	1.000	0.000
Non-prim Dx: acsc_seizure (proportion)	0.005 (0.000)	0.001 (0.000)	0.001 (0.001)	0.000 (0.001)	0.000	0.000	1.000	0.000
Prior Dx: acsc_angina (proportion)	0.003 (0.000)	0.002 (0.000)	0.002 (0.001)	0.000 (0.001)	0.000	0.000	1.000	0.000
Prior Dx: acsc_asthma (proportion)	0.015 (0.001)	0.022 (0.002)	0.023 (0.003)	0.001 (0.005)	4.255	0.007	0.835	0.000
Prior Dx: acsc_copd (proportion)	0.013 (0.001)	0.012 (0.001)	0.012 (0.002)	0.000 (0.003)	0.000	0.000	1.000	0.000
Prior Dx: acsc_dehydration (proportion)	0.010 (0.001)	0.009 (0.001)	0.009 (0.002)	0.000 (0.003)	0.000	0.000	1.000	0.000
Prior Dx: acsc_diab (proportion)	0.065 (0.002)	0.088 (0.004)	0.089 (0.006)	0.001 (0.009)	1.099	0.003	0.913	0.000
Prior Dx: acsc_heartfailure (proportion)	0.009 (0.001)	0.008 (0.001)	0.009 (0.002)	0.000 (0.003)	5.556	0.005	0.857	0.000
Prior Dx: acsc_hypertension (proportion)	0.111 (0.002)	0.117 (0.004)	0.111 (0.007)	-0.006 (0.010)	-5.752	-0.020	0.528	0.000
Prior Dx: acsc_pneumonia (proportion)	0.005 (0.000)	0.003 (0.001)	0.004 (0.001)	0.000 (0.002)	12.500	0.008	0.797	0.000
Prior Dx: acsc_uti (proportion)	0.025 (0.001)	0.014 (0.002)	0.014 (0.003)	0.000 (0.004)	3.448	0.004	0.895	0.000
Prior Dx: acsc_chestpain (proportion)	0.034 (0.001)	0.027 (0.002)	0.027 (0.004)	0.000 (0.005)	0.000	0.000	1.000	0.000
Prior Dx: acsc_cellulitis (proportion)	0.009 (0.001)	0.013 (0.001)	0.013 (0.003)	0.000 (0.004)	0.000	0.000	1.000	0.000
Prior Dx: acsc_dvt (proportion)	0.002 (0.000)	0.005 (0.001)	0.005 (0.002)	0.000 (0.002)	0.000	0.000	1.000	0.000
Prior Dx: acsc_intractpain (proportion)	0.014 (0.001)	0.015 (0.001)	0.016 (0.003)	0.000 (0.004)	3.125	0.004	0.900	0.000
Prior Dx: acsc_sicklecell (proportion)	0.000 (0.000)	0.000 (0.000)	0.001 (0.001)	0.000 (0.001)	50.000	0.016	0.547	0.000

Table A.1 *(continued)*

Measure	Potential comparisons (n = 26740)	Matched comparisons (n = 6145)	Treated (n = 2042)	Adjusted difference (matched)	Percentage difference (matched)	Standardized difference (matched)	*p*-value (matched)	Equivalence *p*-value (matched)
Prior Dx: acsc_syncope (proportion)	0.007 (0.001)	0.003 (0.001)	0.003 (0.001)	0.000 (0.002)	14.286	0.008	0.778	0.000
Prior Dx: acsc_seizure (proportion)	0.015 (0.001)	0.009 (0.001)	0.010 (0.002)	0.001 (0.003)	14.286	0.015	0.621	0.000
Acute hospital admissions	0.640 (0.009)	0.714 (0.017)	0.911 (0.053)	0.197 (0.065)	21.627	0.082	0.002	0.000
Primary care visits - ambulatory	7.014 (0.052)	3.913 (0.068)	3.661 (0.131)	-0.252 (0.175)	-6.886	-0.043	0.149	0.000
Specialist visits - ambulatory	14.577 (0.098)	7.767 (0.117)	7.340 (0.233)	-0.427 (0.296)	-5.818	-0.041	0.149	0.000
Physician services expend	521.527 (4.112)	292.903 (4.357)	289.852 (10.716)	-3.051 (13.481)	-1.053	-0.006	0.821	0.000
Outpatient ED/Obs visits	1.793 (0.011)	2.623 (0.024)	2.892 (0.063)	0.270 (0.087)	9.320	0.095	0.002	0.000
Proportion Medicare/Medicaid dual	0.503 (0.003)	0.535 (0.006)	0.540 (0.011)	0.005 (0.015)	0.917	0.010	0.745	0.000
Race: proportion white	0.530 (0.003)	0.289 (0.006)	0.298 (0.010)	0.009 (0.014)	2.986	0.019	0.526	0.000
Race: proportion black	0.229 (0.003)	0.435 (0.006)	0.429 (0.011)	-0.007 (0.015)	-1.537	-0.013	0.670	0.000
Race: proportion AmIndian Asian/PacIsland	0.134 (0.002)	0.096 (0.004)	0.100 (0.007)	0.004 (0.009)	4.455	0.015	0.632	0.000
Proportion Hispanic	0.088 (0.002)	0.159 (0.005)	0.155 (0.008)	-0.004 (0.011)	-2.676	-0.011	0.717	0.000
Race: proportion unknown	0.019 (0.001)	0.020 (0.002)	0.018 (0.003)	-0.003 (0.004)	-14.954	-0.020	0.532	0.000
Total expenses	2102.889 (22.521)	1869.050 (37.912)	2120.834 (89.891)	251.784 (119.054)	11.872	0.062	0.035	0.000
Outpatient expenses	279.357 (5.472)	316.195 (8.819)	348.026 (15.814)	31.830 (23.196)	9.146	0.045	0.170	0.000
zip code uninsured proportion (proportion)	0.146 (0.000)	0.144 (0.001)	0.144 (0.001)	0.000 (0.002)	-0.074	-0.002	0.951	0.000
hospital occupancy rate	0.816 (0.001)	0.757 (0.002)	0.766 (0.001)	0.009 (0.003)	1.226	0.144	0.004	0.017
Physicians per 100,000	259.576 (0.176)	262.779 (0.341)	261.673 (0.613)	-1.106 (0.878)	-0.423	-0.040	0.208	0.000

Table A.1 *(continued)*

Measure	Potential comparisons (n = 26740)	Matched comparisons (n = 6145)	Treated (n = 2042)	Adjusted difference (matched)	Percentage difference (matched)	Standardized difference (matched)	*p*-value (matched)	Equivalence *p*-value (matched)
PrimDx: Chest pain (proportion)	0.047	0.091	0.091	0.000	0.000	0.000	1.000	0.000
	(0.001)	(0.004)	(0.006)	(0.009)				
PrimDx: Superficial injury (proportion)	0.051	0.016	0.017	0.001	3.088	0.004	0.899	0.000
	(0.001)	(0.002)	(0.003)	(0.004)				
PrimDx: External conditions (proportion)	0.040	0.016	0.015	-0.002	-11.611	-0.014	0.662	0.000
	(0.001)	(0.002)	(0.003)	(0.004)				
PrimDx: Joint disorders, nontrauma (proportion)	0.037	0.033	0.035	0.001	3.756	0.007	0.813	0.000
	(0.001)	(0.003)	(0.004)	(0.006)				
PrimDx: Spondylosis/other back (proportion)	0.032	0.030	0.029	-0.001	-3.389	-0.006	0.848	0.000
	(0.001)	(0.002)	(0.004)	(0.005)				
PrimDx: UTI (proportion)	0.029	0.052	0.052	0.000	0.000	0.000	1.000	0.000
	(0.001)	(0.003)	(0.005)	(0.007)				
PrimDx: Dizziness/vertigo related (proportion)	0.030	0.029	0.034	0.005	15.048	0.028	0.364	0.000
	(0.001)	(0.002)	(0.004)	(0.006)				
PrimDx: Connective tissue (proportion)	0.028	0.040	0.035	-0.005	-13.009	-0.025	0.441	0.000
	(0.001)	(0.003)	(0.004)	(0.006)				
PrimDx: Head/neck wounds (proportion)	0.029	0.005	0.005	0.000	-6.500	-0.005	0.882	0.000
	(0.001)	(0.001)	(0.002)	(0.002)				
PrimDx: Abdominal pain (proportion)	0.028	0.038	0.041	0.003	7.837	0.016	0.597	0.000
	(0.001)	(0.003)	(0.004)	(0.006)				
Proportion male	0.411	0.425	0.417	-0.008	-1.937	-0.016	0.603	0.000
	(0.003)	(0.006)	(0.011)	(0.016)				
Proportion female	0.589	0.575	0.583	0.008	1.384	0.016	0.603	0.000
	(0.003)	(0.006)	(0.011)	(0.016)				
hcc_score	1.853	1.515	1.524	0.010	0.632	0.008	0.803	0.000
	(0.009)	(0.015)	(0.028)	(0.039)				
propensity score	0.059	0.197	0.230	0.033	14.515	0.184	0.000	0.010
	(0.001)	(0.001)	(0.004)	(0.005)				

	Chi-squared statistic	Degrees of freedom	*p*-value
Omnibus test for balance on matching variables	174.632	37.000	0.000

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data.

Note: Standard errors in parentheses. Standardized difference calculated as the difference in means divided by the treatment group standard deviation. *p*-values come from a weighted two-sample t-test. The equivalence test comprises two one-sided tests of whether the difference between the treatment and comparison group means exceeds 0.25 standard deviations (in either direction) of that variable. If we can reject that hypothesis, we are confident that the true difference between the two populations being compared is less than 0.25 standard deviations.

COPD = chronic obstructive pulmonary disease; HCC = hierarchical condition code; UTI = urinary tract infection.

This page has been left blank for double-sided copying.

APPENDIX B

IMPACT ESTIMATE TABLES

This page has been left blank for double-sided copying.

Table B.1. Estimated impact of the New York City Health + Hospitals program on average Medicare FFS expenditures (PBPM) during a 12-month follow-up period

	Baseline period	Intervention period	Difference (intervention minus baseline)	*p*-value	90% confidence interval
Total[a]					
Comparison beneficiaries	1,645	2,175	531		
Treatment beneficiaries	2,040	2,637	597		
Difference (treatment minus comparison)	395	461	66	0.592	(-137, 270)
Acute inpatient					
Comparison beneficiaries	731	1,017	286		
Treatment beneficiaries	1,144	1,484	340		
Difference (treatment minus comparison)	413	467	54	0.557	(-97, 205)
Other inpatient					
Comparison beneficiaries	122	144	22		
Treatment beneficiaries	86	113	27		
Difference (treatment minus comparison)	-37	-31	6	0.829	(-38, 49)
Outpatient					
Comparison beneficiaries	302	310	8		
Treatment beneficiaries	342	361	20		
Difference (treatment minus comparison)	39	52	12	0.465	(-15, 40)
Professional					
Comparison beneficiaries	281	369	89		
Treatment beneficiaries	287	339	52		
Difference (treatment minus comparison)	7	-30	-37***	0.004	(-58, -16)
Durable medical equipment					
Comparison beneficiaries	17	19	2		
Treatment beneficiaries	13	14	0		
Difference (treatment minus comparison)	-4	-5	-1	0.738	(-8, 5)
Home health					
Comparison beneficiaries	80	95	15		
Treatment beneficiaries	58	72	14		
Difference (treatment minus comparison)	-22	-23	-1	0.861	(-12, 9)
Skilled nursing facility					
Comparison beneficiaries	113	238	129		
Treatment beneficiaries	109	206	92		
Difference (treatment minus comparison)	-4	33	37	0.214	(-11, 96)
Hospice					
Comparison beneficiaries	-2	15	17		
Treatment beneficiaries	1	15	4		
Difference (treatment minus comparison)	3	0	-3	0.568	(-12, 6)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment data through September 30, 2017.

Note: Number of treatment beneficiaries = 1,589; number of comparison beneficiaries = 4,778.

[a]Total Medicare expenditures were calculated from all claims for each participant with at least one eligible day during the baseline year and quarter.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

FFS = fee-for-service; PBPM = per beneficiary per month.

Table B.2. Estimated impact of the New York City Health + Hospitals program on service utilization rates (per 1,000 beneficiaries) during a 12-month follow-up period

	Baseline period	Intervention period	Difference (intervention minus baseline)	*p*-value	90% confidence interval
Acute hospital admissions[a]					
Comparison beneficiaries	353	507	154		
Treatment beneficiaries	865	1,048	183		
Difference (Treatment minus comparison)	512	541	29	0.618	(-68, 126)
Outpatient ED visits[b]					
Comparison beneficiaries	3,218	2,666	-552		
Treatment beneficiaries	3,610	3,086	-524		
Difference (Treatment minus comparison)	393	420	28	0.894	(-313, 368)
Primary care visits in ambulatory settings					
Comparison beneficiaries	4,623	5,832	1,209		
Treatment beneficiaries	3,665	4,251	586		
Difference (Treatment minus comparison)	-958	-1,581	-623***	0.001	(-926, -321)
Specialist visits in ambulatory settings					
Comparison beneficiaries	8,986	11,009	2,022		
Treatment beneficiaries	7,415	8,417	1,002		
Difference (Treatment minus comparison)	-1,572	-2,592	-1,020***	0.005	(-2,358, -605)
Specialist visits in any setting					
Comparison beneficiaries	12,952	16,392	3,440		
Treatment beneficiaries	11,240	13,199	1,959		
Difference (Treatment minus comparison)	-1,712	-3,194	-1,481***	0.005	(-1,616, -424)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment data through September 30, 2017.

Note: Number of treatment beneficiaries = 1,589; number of comparison beneficiaries = 4,778.

[a]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[b]Includes visits to an ED, as well as observation stays.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department.

Table B.3. Estimated impact of the New York City Health + Hospitals program on percentage of any acute care utilization during a 12-month follow-up period

	Baseline period	Intervention period	Difference (intervention minus baseline)	*p*-value	90% confidence interval
Percentage with any acute hospital admissions[a]					
Comparison beneficiaries	20.8	24.8	4		
Treatment beneficiaries	22.6	25.7	3.1		
Difference (treatment minus comparison)	1.8	0.9	-0.9	0.448	(-2.8, 1)
Percentage with any outpatient ED visits[b]					
Comparison beneficiaries	67.8	39.9	-27.9		
Treatment beneficiaries	69.9	44.2	-25.7		
Difference (treatment minus comparison)	2.1	4.3	2.2*	0.075	(0.2, 4.3)
Percentage with any 30-day readmission among all participants					
Comparison beneficiaries	2.5	5.3	2.8		
Treatment beneficiaries	5.5	7.3	1.8		
Difference (treatment minus comparison)	3.0	2.0	-1.0	0.157	(-2.1, 0.2)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment data through September 30, 2017.

Note: Number of treatment beneficiaries = 1,589; number of comparison beneficiaries = 4,778.

[a]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[b]Includes visits to an ED, as well as observation stays.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department.

Table B.4. Estimated impact of the New York City Health + Hospitals program on average Medicare FFS expenditures (PBPM) during a 24-month follow-up period

	Baseline period	Intervention period	Difference (intervention minus baseline)	*p*-value	90% confidence interval
Total[a]					
Comparison beneficiaries	1,702	2,219	517		
Treatment beneficiaries	2,040	2,503	463		
Difference (treatment minus comparison)	338	284	-54	0.683	(-272, 164)
Acute inpatient					
Comparison beneficiaries	691	994	303		
Treatment beneficiaries	1,144	1,388	244		
Difference (treatment minus comparison)	453	394	-59	0.553	(-222, 104)
Other inpatient					
Comparison beneficiaries	153	153	0		
Treatment beneficiaries	86	96	10		
Difference (treatment minus comparison)	-67	-5	10	0.670	(-29, 49)
Outpatient					
Comparison beneficiaries	337	342	4		
Treatment beneficiaries	342	368	27		
Difference (treatment minus comparison)	4	27	22	0.345	(-17, 61)
Professional					
Comparison beneficiaries	294	382	88		
Treatment beneficiaries	287	340	53		
Difference (treatment minus comparison)	-7	-41	-35	0.029**	(-61, -9)
Durable medical equipment					
Comparison beneficiaries	19	19	0		
Treatment beneficiaries	13	16	3		
Difference (treatment minus comparison)	-5	-3	2	0.598	(-5, 9)
Home health					
Comparison beneficiaries	84	101	18		
Treatment beneficiaries	58	61	3		
Difference (treatment minus comparison)	-26	-40	-15	0.071	(28, -1)
Skilled nursing facility					
Comparison beneficiaries	128	212	84		
Treatment beneficiaries	109	220	111		
Difference (treatment minus comparison)	-19	8	27	0.309	(-17, 70)
Hospice					
Comparison beneficiaries	-4	17	20		
Treatment beneficiaries	1	14	13		
Difference (treatment minus comparison)	5	-3	-8	0.598	(-19, 4)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment data through September 30, 2017.

Note: Number of treatment beneficiaries = 551; Number of comparison beneficiaries = 1,782.

[a]Total Medicare expenditures were calculated from all claims for each participant with at least one eligible day during the baseline year and quarter.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

FFS = fee-for-service; PBPM = per beneficiary per month.

Table B.5. Estimated impact of the New York City Health + Hospitals program on service utilization rates (per 1,000 beneficiaries) during a 24-month follow-up period

	Baseline period	Intervention period	Difference (intervention minus baseline)	*p*-value	90% confidence interval
Acute hospital admissions[a]					
Comparison beneficiaries	322	467	145		
Treatment beneficiaries	865	1,029	164		
Difference (treatment minus comparison)	543	562	19	0.772	(-89, 127)
Outpatient ED visits[b]					
Comparison beneficiaries	3,616	3,320	-296		
Treatment beneficiaries	3,610	2,802	-808		
Difference (treatment minus comparison)	-6	-518	-512	0.233	(-1,219, 194)
Primary care visits in ambulatory settings					
Comparison beneficiaries	4,538	5,874	1,336		
Treatment beneficiaries	3,665	4,314	649		
Difference (treatment minus comparison)	-873	-1,560	-687***	0.003	(-1,063, -313)
Specialist visits in ambulatory settings					
Comparison beneficiaries	9,181	11,273	2,092		
Treatment beneficiaries	7,415	8,249	834		
Difference (treatment minus comparison)	-1,766	-3,024	-1,258**	0.018	(-2,136, -380)
Specialist visits in any setting					
Comparison beneficiaries	13,706	17,423	3,717		
Treatment beneficiaries	11,240	13,180	1,940		
Difference (treatment minus comparison)	-2,466	-4,243	-1,777**	0.047	(-3,250, -305)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment through of September 30, 2017.

Note: Number of treatment beneficiaries = 551; Number of comparison beneficiaries = 1,782.

[a]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[b]Includes visits to an ED, as well as observation stays.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department.

Table B.6. Estimated impact of the New York City Health + Hospitals program on percentage of any acute care utilization during a 24-month follow-up period

	Baseline period	Intervention period	Difference (intervention minus baseline)	*p*-value	90% confidence interval
Percentage with any acute hospital admissions[a]					
Comparison beneficiaries	20.0	23.4	3.4		
Treatment beneficiaries	22.6	25.2	2.6		
Difference (Treatment minus comparison)	2.6	1.8	-0.08	0.484	(-2.8, 1.1)
Percentage with any outpatient ED visits[b]					
Comparison beneficiaries	67.7	37.7	-30		
Treatment beneficiaries	69.9	42	-27.9		
Difference (Treatment minus comparison)	2.2	4.3*	2.1*	0.090	(0.1, 4.1)
Percentage with any 30-day readmission among all participants					
Comparison beneficiaries	2.1	4.7	2.6		
Treatment beneficiaries	5.5	6.7	1.2		
Difference (Treatment minus comparison)	3.4	2	-1.4*	0.073	(-2.6, -0.1)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment data through September 30, 2017.

Note: Number of treatment beneficiaries = 551; Number of comparison beneficiaries = 1,782.

[a]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[b]Includes visits to an ED, as well as observation stays.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department.

This page has been left blank for double-sided copying.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE,WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Seattle Children's Hospital

April 18, 2018

Ella Douglas-Durham (Mathematica Policy Research)
Victoria Peebles (Mathematica Policy Research)
Marisa Morrison (RTI)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Seattle Children's Hospital received a no-cost extension through May 31, 2018. The awardee stopped providing Pediatric Partners in Care (PPIC) services after August 31, 2017. During the no-cost extension, the awardee planned to focus on using data to further drive ongoing payment model negotiations.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

Seattle Children's Hospital, a tertiary care medical center, used its HCIA R2 funds to implement the Pediatric Partners in Care (PPIC) program for children with medical complexity who were enrolled in both Medicaid and the Supplemental Security Income (SSI) programs. The awardee sought to provide care coordination across hospital, primary care, and community settings and work closely with Medicaid managed care organizations (MCOs) to develop a sustainable service delivery and payment model. The goals of the program were to (1) improve the health outcomes of children with disabilities who have Medicaid and SSI coverage; (2) reduce medical costs for these children by eliminating unnecessary, redundant, and ineffective treatments; and (3) develop a scalable management model for outpatient care that supports and optimizes the existing infrastructure for delivering care. In addition to managing and coordinating care, Seattle Children's Hospital also trained the children's primary care providers and school nurses in managing specific conditions. For example, the awardee developed and presented tools that guide providers in making decisions about and managing feeding tubes. The program was a new approach in the region toward coordinating care for children with medical complexity and was innovative in its tight relationships with Medicaid MCOs in developing services and the payment model.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Seattle Children's Hospital's PPIC program aims to (1) improve the health outcomes of children with disabilities who have SSI and Medicaid coverage; (2) reduce medical costs for these children by eliminating unnecessary, redundant, and ineffective treatments; and (3) develop a scalable management model for outpatient care that supports and optimizes the existing care delivery infrastructure.
Major innovation	Relationships with MCOs in developing intervention services and payment model
Program components	• Care coordination and management • Provider education and training
Target population	Children and adolescents younger than 18 who: • Live in King and Snohomish counties in WA • Are enrolled in the SSI program • Are covered by Medicaid and enrolled in one of four MCOs • Are identified as being at high risk for negative health outcomes
Payment model	Per beneficiary per month (PBPM) care management fee adjusted for measures of quality, utilization, and spending
Award amount	$5,561,620
Effective launch date	2/1/2015
Program setting	Provider-based
Market area	Urban, suburban
Market location	King and Snohomish counties
Target outcomes	• Improve measures of care coordination by 10% for the majority of participants • Improve measures of a child's quality of life by 10% for half of the participants • Reduce the overall cost of care by 9.7%

MCO = managed care organization; PPIC = Pediatric Partners in Care; SSI = Supplemental Security Income.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, the awardee nearly met its goal for number of participants, directly serving 813 children—85 percent of its target—by the end of the initial cooperative agreement. Second, the awardee exceeded its goals for enrolling primary care practices, enrolling 34 practices. Third, the program was successful in fully staffing the PPIC program. Fourth, awardee-reported data and survey data suggest the program successfully engaged participants. Finally, participating clinicians and other implementation staff reported their perceptions that the program had a positive effect on the delivery of care.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Seattle Children's Hospital's program, the analysis is still in progress and not included in this report.

Payment model. Seattle Children's Hospital worked with Medicaid MCOs to develop a framework for a PBPM care management fee that would be adjusted based on measures of quality, utilization, and spending. For the final year of the cooperative agreement, the awardee had contractual agreements with four MCOs for testing and developing the model without payment, and planned to negotiate payment contracts based on findings from the test period.

Sustainability plans. The awardee reported plans to only sustain parts of the program it believes have been effective, and did not plan to sustain the intervention in its original form. Seattle Children's Hospital will continue to develop the payment model and engage payers during the no-cost extension period, which ends in May 2018. The awardee did not report plans to scale or replicate its program without a successful care management contract with one or more MCOs.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July 2016 to October 2016 with all program staff members at the time (nine), and had a response rate of 89 percent. The clinician survey was fielded from March 2017 to June 2017 with 11 clinicians, and had a response rate of 100 percent. Two of the respondents of the clinician survey were also program leaders for the PPIC program, and the remaining clinicians were primary care physicians (PCPs) of enrolled beneficiaries and/or involved in the provider education and training component of the project. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a] Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

PPIC targeted about 4,000 children and adolescents younger than 18 in King and Snohomish counties in Washington State who were enrolled in the SSI program, had Medicaid coverage, and were enrolled in one of four MCOs. Of these 4,000 children, PPIC staff aimed to coordinate care directly for 960 children who were identified as being at high risk for negative health outcomes. A child's risk was determined on the basis of the following factors: a hospitalization within the past six months, two ED visits in six months, or a Washington State Predictive Risk Intelligence

System (PRISM) score greater than 1.[4] The program staff also enrolled any PPIC-eligible children who receive care from a PCP who was engaging with the program.

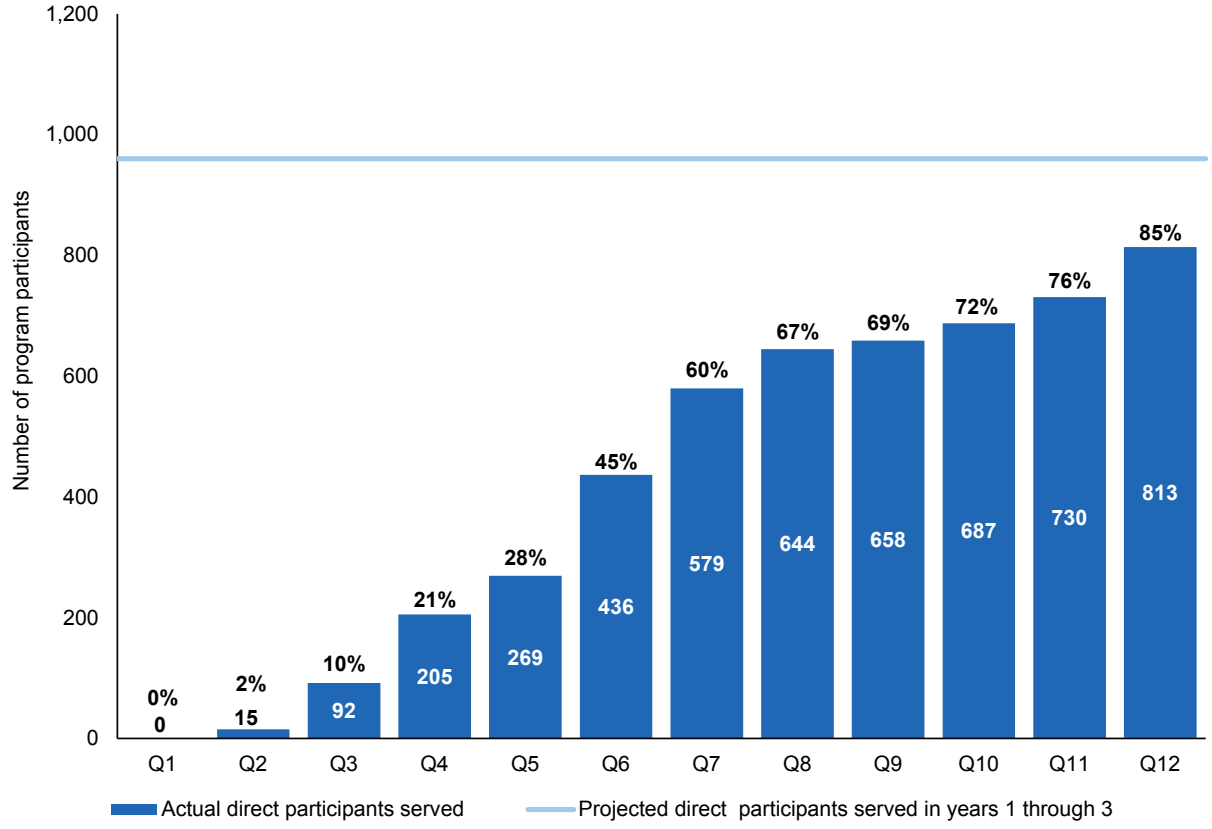
Seattle Children's Hospital experienced delays enrolling participants, due primarily to delays receiving data from each MCO and the time required to contact and enroll participants. First, because Seattle Children's Hospital depended upon Medicaid MCO data to identify children for enrollment, the awardee could not start enrollment until February 2015 (five months into the cooperative agreement). This was due to the time needed to negotiate with MCOs for access to data, data files that were not usable initially, and the challenge of aggregating files across MCOs. Second, the awardee noted that enrollment continued to lag until the third quarter of the program due to the time required for staff to contact caregivers and complete an initial assessment, which was necessary before offering the child enrollment. To mitigate this challenge, and in response to changes in the reporting that distinguished "indirect" from "direct" intervention, during the third quarter of the program Seattle Children's Hospital revised its enrollment process to passively enroll indirect participants, who became the eligible pool for direct participants. Program staff continued to use claims data to identify children eligible to participate in PPIC and sent a letter to the child's caregiver welcoming them into the program. Children were considered enrolled as indirect participants if the caregiver did not opt out, and only became direct participants if the caregiver completed the assessment process and opted in. Seattle Children's Hospital changed its projections from the beginning of the first year (when it projected 1,600 direct and 162 indirect participants) to the beginning of the third year (960 direct and 1,600 indirect participants), to align with the reporting change. It reduced the projected number of direct participants who would receive care management services and increased the number of indirect participants due to the estimated spillover influences of the provider education component and monthly care management meetings with MCOs. (Indirect participants refers to the unique participants who received program services through education and training provided to PCPs.)

b. Evidence of enrollment effectiveness

Overall, the awardee reported that it enrolled 813 participants from February 2015 (when it launched its program) through August 2017, which represents about 85 percent of its 960 three-year projected direct participants (Figure II.1). We concluded that Seattle Children's Hospital was partially effective in enrollment because it did not reach 90 percent of its three-year enrollment target by the end of the 12th program quarter despite revising its original enrollment goal downward. The awardee lowered its direct participant enrollment projection from 1,600 at the beginning of Year 1 to 960 at the beginning of Year 2, in response to changes in reporting to define direct participants only as those for whom the caregiver completed the assessment process and opted into the program. The enrollment projection stayed the same at the beginning of Year 3.

[4] A predictor of health care utilization developed by Washington State's Medicaid agency based on the Chronic Illness and Disability Payment System for Medicaid.

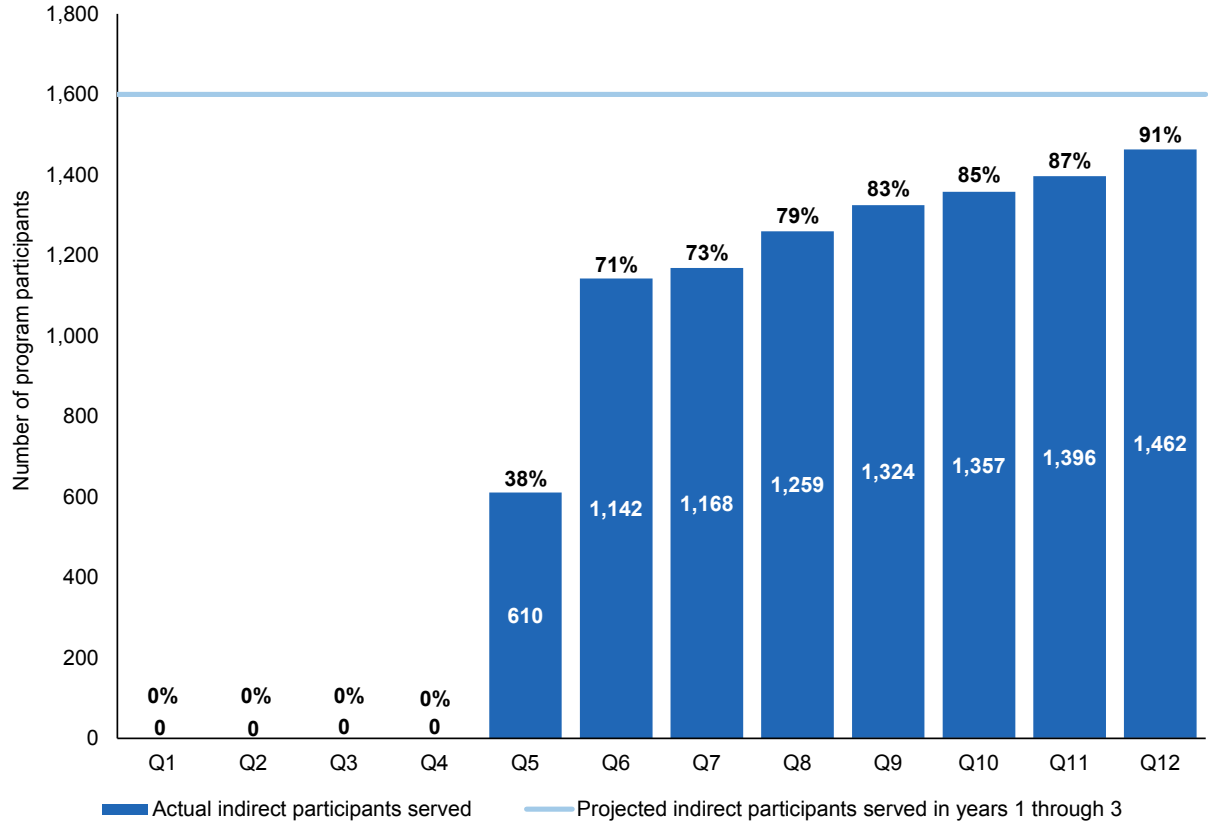
Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee lowered its direct participant enrollment projection from 1,600 at the beginning of Year 1 to 960 at the beginning of Year 2, in response to changes in reporting to define direct participants only as those for whom the caregiver completed the assessment process and opted into the program. The enrollment projection stayed the same at the beginning of Year 3.

Figure II.2. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers.

Seattle Children's Hospital also reported that it enrolled 1,462 indirect participants from February 2015 through August 2017, which represents 91 percent of its 1,600 three-year projected indirect participants (Figure II.2).

c. Barriers and facilitators associated with enrollment effectiveness

Seattle Children's Hospital's progress in meeting its three-year enrollment goals for both participants and providers was influenced by several factors, including changes to the program's enrollment approach and children's SSI status.

Participant recruitment. First, Seattle Children's Hospital changed its enrollment processes throughout the life of the PPIC program, which facilitated program enrollment. During the first program year, the awardee added a passive enrollment approach for indirect participants. This change enabled Seattle Children's Hospital to count children as enrolled in the indirect intervention as soon as they received a letter welcoming them into the program, removing the challenge of waiting for the care team to reach out to the beneficiary and helping to enroll more

children faster. In the third program year, the awardee evaluated its program and again refined its enrollment process to target specific populations (for example, beneficiaries with asthma, who were also enrolled in Medicaid and SSI). Based on lessons learned early on in the program, the program focused on specific populations for which it could make an impact in a shorter period, as leaders identified reduced effectiveness with longer term enrollees, and as the program was ending, it was important to be clear that the intervention period was more limited for new enrollees.

> "It's taken about the first two years, I would say, to really get an understanding of what services we can provide. In Year 3, it's been about, 'Oh, you can provide those services. That's great. I need that help and my friend down the street at the other clinic may need that help.'"
>
> —Care manager

Second, participants' SSI status was a challenge for enrollment. During the first program year, program staff identified a number of children who were eligible for SSI and PPIC based on all other criteria, but who were not enrolled in SSI. As PPIC staff began working with providers for children enrolled in the program, PCPs also identified other children in their practices who would be eligible for the program but were not enrolled in SSI. As a result, PPIC staff began to help families in participating practices to enroll in SSI and in PPIC. According to one interview respondent, "When we identify a patient who has a need [for care coordination], it may take us five months to even get them on it, to get them eligible for the program. That has been a challenge because oftentimes the providers are identifying patients who don't have SSI and therefore don't qualify for our services." In addition, children's enrollment in SSI could quickly shift, which required staff's assistance re-enrolling the child to receive SSI benefits so he or she could maintain PPIC eligibility. As one respondent noted, "SSI is a very transient service and support for families, and so they may come on it and be off it within six months."

Third, staff noted a decreased emphasis on enrollment and the intake process in the third program year. Program staff reported that other priorities, such as time training new staff or planning for an asthma training event, took staff time away from enrolling participants directly. One care manager indicated that the lack of emphasis on intakes and enrollment as a priority in the later years of the program made it harder to meet enrollment goals.

Provider recruitment. PPIC staff indicated they initially recruited a group of PCPs who recognized and appreciated Seattle Children Hospital's effort to fill gaps in care and connect and coordinate existing resources within the community to improve care coordination. Interview respondents also indicated that the PPIC program's affiliation with Seattle Children's Hospital, a large, well-known hospital in the region, facilitated provider recruitment. The program then had success recruiting additional primary care practices through providers who were already connected to the program. PCPs who were engaged with the PPIC program connected program staff to PCPs at other primary care practices who they knew could benefit from the PPIC program.

One barrier to recruiting primary care practices was in demonstrating the advantage of pediatric-specific care management in the PPIC program to practices with existing care coordination resources. One respondent noted that some providers viewed the PPIC program as just another external program, and they did not see an incentive to participating, particularly if their practice had an existing care management program. This barrier was a particular issue at

Federally Qualified Health Centers, as they have several existing care management and coordination resources.

2. Delivery of program services

a. Description of and changes to service delivery model

The PPIC program was a new program at Seattle Children's Hospital, which created four care teams that provided comprehensive care management and coordination services. Each team, which consisted of a nurse care manager and a community care coordinator, engaged participants' families through an initial PPIC assessment to identify barriers to and gaps in the children's care. The care team then developed a care plan in collaboration with the family, and the nurse care managers held regular phone calls with the family to review the care plan and reassess participants' needs. The calls took place weekly or monthly depending on the family's level of need. In addition, the nurse care managers held regular calls with the child's PCP or the PCP's staff. The community care coordinator served as a community health worker and care navigator who helped families connect with community resources and navigate the health care, school, and social services systems.

Seattle Children's Hospital documents workflows (called *process pages*) and has adapted them over the course of the program as the interventions evolved or gaps in processes were identified. In the third year, Seattle Children's Hospital developed process pages for graduating patients out of the PPIC program, integrating the health home model into the program, and creating the maintenance program. In fall 2016, the program began trialing different programs such as a home visiting program for asthma patients and a Health Homes program, and conducting home visits after a transition of care.

Participants enrolled in the beginning of the program received services for as long as they were enrolled in the program. However, during Year 3, the program began "graduating participants" if the family did not identify any outstanding needs during a scheduled call with the nurse care manager. PPIC staff provided information to graduated families on where they should call for needs in the future, and gave them the option to reenroll if they identified additional needs the program could address. In Year 3 the program also focused more on providing shorter-term interventions, including a home visiting program for asthma patients and a 30- to 45-day intervention after a transition of care.

In addition to enrolling individual children, PPIC staff also identified practices with multiple children enrolled in the program and worked with these practices to identify opportunities for education to improve care. At the start, the program only had a minor focus on engaging primary care practices, but PPIC leaders reported that the first practices they engaged had been eager to "actively participate in system improvements such as inpatient discharge communication and hand-off, challenging medication regimens, and marketing of urgent care options." As a result, beginning in the fourth quarter of the cooperative agreement (June 2015–August 2015), program staff began to spend more time identifying the strengths and needs among community primary care practices that provided care to participants. Beginning in Year 2, the program also began to reach out to school nurses in the community and held targeted training events on asthma and feeding tubes for this group.

b. Evidence of service delivery effectiveness

After initial delays in enrollment, Seattle Children's Hospital was largely successful in implementing its service delivery model. The awardee met its goals for staffing and training, recruiting providers, and engaging participants, but experienced delays in providing services. The awardee adapted the program's focus over the course of the cooperative agreement in an effort to make iterative improvements in service delivery and to target the interventions to where they might have the largest short-term impact.

Delivering intervention services. The program experienced delays in providing services until the second year of the program, but the interviews and responses from the staff surveys indicated that the program was successful in delivering services to children and their families as intended after that time. The focuses of the intervention evolved over the course of the cooperative agreement. Awardee leaders noted that these changes were driven by their use of quality improvement cycles to identify opportunities for improvement and implement interventions that were more likely to help them reach their goals.

Findings from the staff survey demonstrated that most staff were engaged in the types of activities that the program intended. Examples included educating participants about managing their own care and services; assisting participants with accessing non-medical services such as housing, job training, or supplemental nutrition services (for example, Supplemental Nutrition Assistance Program benefits); attending medical appointments with participants; conducting home visits; and coaching participants.

Staffing and training. After overcoming initial staff recruitment delays, Seattle Children's Hospital successfully hired and retained the staff it needed to deliver services as intended for the PPIC program. Based on data from the implementation and monitoring contractor, Seattle Children's Hospital hired 10.49 full-time equivalent staff since project inception, which is 116.6 percent of its three-year hiring projection. The awardee hired staff for four care teams (four nurse care managers and four community health coordinators), and leaders reported that they had no issues hiring care managers or community care coordinators. It also hired a program coordinator. In addition, the awardee used existing staff, which included a research coordinator and data analyst to work on Medicaid data, and a medical director who coordinated with PCP practices and provided clinical advice to the care team.

Staff turnover was also not a reported problem. The awardee reported to the implementation and monitoring contractor that as of August 2017, it had achieved a 100 percent retention rate for staff. Seattle Children's Hospital leaders reported that they kept staff engaged by asking them to analyze what worked with their population and adapting the intervention to best suit the needs of the population, especially through creating the home visiting program for asthma patients and the intervention for care transitions.

According to information from the interviews and survey data, staff perceived their training as effective. In site visit interviews, care team staff reported that the variety of prior experiences from staff members was helpful. Interview respondents felt that these different backgrounds strengthened the team and enabled staff to train one another. For example, one staff person shared her knowledge about Medicaid from a previous role where she enrolled families in Medicaid.

Recruiting and engaging providers. The site visit interviews and the awardee's operational protocol indicated that the awardee reported enrolling 34 PCPs, exceeding its target of 20 PCPs. Nearly all clinicians who completed the survey reported receiving formal training, and most strongly agreed that it helped to improve their job performance.

Engaging program participants. According to awardee-reported data and survey findings, PPIC successfully engaged participants and their families by providing care management and care coordination services. All of the staff surveyed at the end of the second program year either somewhat agreed or strongly agreed that they had successfully engaged participants with the program. The majority of the clinicians surveyed felt that participant engagement was the most helpful to PPIC achieving its goals, and strongly agreed that they had engaged participants in this program. Based on responses to its own survey of caregivers of about 150 participants after 12 months and 80 participants after 18 months, the awardee reported increases compared with baseline in the proportion of caregivers who reported the following measures: (1) that their child has a designated care coordinator, (2) that their care coordinator helped their child to obtain needed community resources, (3) the care coordinator has contacted them or attempted to contact them at least once in the last three months, and (4) the care coordinator advocated for the needs of the child. The awardee also reported the proportion of caregivers who reported that their child had a shared care plan. This measure did not improve over time, staying consistent at around one-quarter of respondents. Program staff reported that the software used to create care plans does not have an export function, which remained a challenge throughout the duration of the program.

c. Barriers and facilitators associated with effectiveness of service delivery

In prior years, the awardee faced delays in receiving enrollment data, hiring staff, and enrolling participants, but by the third year of the cooperative agreement, these challenges were no longer an issue. In the third year, awardee and site leaders focused on adapting the program to best meet the needs of participants and continuing to engage providers.

In interviews, program and site leaders reported that, due to delays in receiving enrollment data and staffing the program, as well as a time-intensive enrollment process, the program was not fully delivering services as intended until its second year. Delays in hiring staff occurred for a number of reasons. First, Seattle Children's Hospital's leaders noted that receiving the HCIA funding notification immediately preceding the start date was a challenge for an organization in which several months elapse from the time a position is defined to when it can be filled. Second, the awardee also experienced initial challenges in hiring care managers who are comfortable with coordinating outpatient care rather than inpatient care, and with recruiting community outreach workers with the desired skill set. By August 2015 (almost a year after the beginning of the cooperative agreement), Seattle Children's Hospital had hired two nurse care managers, and by September 2015 it had hired two community outreach workers.

Another facilitator highlighted in the third year was maintaining PCP engagement by demonstrating the advantage of the PPIC program. Care team staff maintained engagement in a number of ways. First, they kept PCPs up to date and engaged through phone calls to discuss participants and to update them on programmatic changes so the PCPs remain active participants in the program. Second, PPIC staff made themselves accessible to the PCPs when they had

needs, questions, or concerns. Third, program staff assisted PCPs with patient referrals, when needed. Care team staff have taken advantage of their connection with Seattle Children's Hospital to access resources quicker and easier.

C. Assessment of perceived program effects on the delivery of care and outcomes

The majority of program staff and clinicians who responded to surveys perceived that the PPIC program had positive impacts on care coordination and delivery. An overwhelming majority of respondents across both surveys said the PPIC program had a positive impact on the quality of care and services they provided to patients and that the program had a positive impact on patient quality of life. Survey respondents also strongly believed that the program had a positive impact on patient satisfaction.

Interview respondents echoed the survey findings regarding care coordination. Frontline staff expressed they had success helping coordinate care across the continuum of care, and program leaders believed they were already beginning to see cost savings as a result of the program, especially related to reductions in redundant specialty visits and proactive care management. In addition, program leaders reported that their survey of families' experience with care coordination showed improvement in the proportion of families that identified having a single care coordinator arrange their care, that the care coordinator helped their child to obtain needed community services, and that their child's care coordinator advocated for the needs of their child. Frontline staff also noted that they have had success in educating patient families on appropriate ED use and that patient families have demonstrated a better understanding on using the ED versus urgent care, or have called the care team if they have a concern, rather than immediately going to the ED.

Clinicians completing the clinician survey indicated a number of factors were helpful to the success of the PPIC program. A majority of clinicians said the PPIC program has made their job easier overall. Most clinicians indicated that fellow clinicians' buy-in to the program was helpful to the program achieving its goals. Other factors identified were sufficient staffing, patient engagement in the program, and effective communication. Interview respondents and site-level staff echoed some of these findings; interview respondents felt that the targeted interventions later in the cooperative agreement, such as the home visiting program for asthma patients, greatly promoted patient engagement in the program by addressing specific patient and family needs.

Despite the overwhelming agreement among survey and interview respondents that they perceived positive effects resulting from the PPIC program, interview respondents noted that for some outcomes, such as utilization and cost of care, the PPIC program might not make an impact. This was not because the program was unnecessary or unsuccessful, but because for some patients care coordination will not affect their utilization or the overall cost of care. For example, one care manager indicated that educating a patient with cystic fibrosis about her medications or disease process will not necessarily reduce the child's chances of going to the ED because there still might be instances when it is necessary for the child to go to the ED. Another care manager provided a similar example, noting that it is often appropriate for high-risk asthma patients to visit the ED when they are experiencing respiratory distress, and care coordination

will not reduce these types of visits. Other interview respondents noted that it might take longer than the length of the cooperative agreement to see an impact on cost of care.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. First, the program was not fully ramped up until its second year, so we did not expect any impacts in the first year. Second, the focus of the program evolved from broad care coordination during the second year to more focused, condition-specific interventions during the third year. Impact findings from the second year of the intervention are likely to reflect the broad care coordination, whereas any findings from the third year would reflect the transition to the more-focused interventions. Third, several of the awardee's targeted outcomes are not measurable in Medicaid administrative and claims data, which we plan to use in our impact evaluation, specifically measures of care coordination and quality of life. Lastly, the program could have improved on those goals without decreasing health care spending if any reductions in hospitalizations or redundant services were offset by increased use of needed and recommended services, such as preventive medications and follow-up visits to monitor conditions.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Seattle Children Hospital's PPIC program.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Seattle Children's Hospital

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018	4,943[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	2,932
Likelihood of all-cause hospitalizations	380
MDE sample size requirement to detect 20% effect	
Total expenditures	733
Likelihood of all-cause hospitalizations	95
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	No serious issues; proceeding with comparison group selection
Do claims identify the primary expected effects?	Some effects observed in claims data, but important effects likely missing
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician, staff, and beneficiary surveys
Implementation data that will be analyzed	Awardee is analyzing its own caregiver survey

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We anticipate conducting a rigorous impact analysis of the PPIC program, though we have not yet started an analysis because of the lag in obtaining Medicaid data. The projected size of the analysis sample shown in Table III.1 (based on the date enrollment ended, August 31, 2017) suggests that the sample should be large enough to detect plausible impacts. The criteria for selecting the treatment group are well defined and could be applied to claims data from other Medicaid beneficiaries in the state to form a potential comparison group. We will then use propensity score matching to select comparison beneficiaries whose characteristics and prior Medicaid service use are similar to those of beneficiaries in the treatment group.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Seattle Children's Hospital worked with Medicaid MCOs to develop a framework for a PBPM care management fee that would be adjusted based on measures of quality, utilization, and spending. For the final year of the cooperative agreement, the awardee had contractual agreements with four MCOs for testing and developing the model without payment, and planned to negotiate payment contracts based on those findings.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Seattle Children's Hospital engaged with Medicaid MCOs to develop a new payment model for the target population of the cooperative agreement—children enrolled in SSI and Medicaid in King and Snohomish counties in Washington State. The model is specific to the care coordination and management services developed under the cooperative agreement and does not include primary care practices or individual clinicians who deliver care to the target population. Seattle Children's Hospital worked with four of five MCOs in the state to agree to a standard framework for the payment model.

Under the model, MCOs contract with the awardee to provide a PBPM care management fee for each child in the MCO enrolled in SSI. In each new year of the contract, the fee is adjusted based on weighted performance on: (1) average spending for the enrolled population compared to a pre-intervention baseline (35 percent of fee adjustment); (2) utilization rates (hospitalization, readmission, and ED use) compared to a pre-intervention baseline (40 percent of fee adjustment); and (3) performance on a set of program-specific process measures (25 percent of fee adjustment). The specific amount of the fee will be negotiated between Seattle Children's Hospital and each MCO; as of September 2017, they had not yet determined this fee.

C. Status of the payment model

The awardee was continuing to negotiate with MCOs about the model at the end of the third year of the cooperative agreement. Seattle Children's Hospital reached contractual agreements with the four MCOs during the third year for a "model" year, during which the MCOs would provide claims data to the awardee and all of the MCOs would assess the financial impacts of the model had it been fully implemented. The awardee and MCOs continued to meet regularly to discuss results and potential adjustments to the model. In addition, the awardee signed contracts with three of the MCOs during the third year of the program to serve as the provider for children in the MCOs eligible for the state's Medicaid Health Homes program. The awardee leaders described that their primary motivation for participating in the Health Homes program was to continue to build on their relationships with MCOs and to compare the PPIC intervention with the Health Homes model. Leaders in the program noted that the Health Homes program only

applies to about 30 percent of the participants in PPIC because of differing eligibility, and the monthly fees the Health Homes program pays falls well below the costs of the program for enrolling and engaging with families of eligible children. As a result of cooperation during the program, Seattle Children's Hospital has also been in discussions with at least one of the MCOs about a broader value-based payment model.

> "We're out of time. We just have three quarters of [cost data] and that's not enough for [payers] to feel comfortable."
>
> —Program leader

D. Factors associated with the development of the payment model

In awardee documents and interviews, awardee leaders identified two primary facilitators to developing and negotiating the payment model: (1) building relationships with payers and (2) agreeing to a preliminary testing stage after which the model would be refined. Program leaders from Seattle Children's Hospital described relationship building as pivotal in the awardee's success in bringing together all four MCOs. Leaders at Seattle Children's Hospital met with MCO executives before program launch, and they continued to meet at least quarterly. These meetings included the finance and actuarial leaders at the MCOs, whom awardee leaders described as important in designing and refining the payment model in a form attractive to the MCOs. The awardee's actuarial consultants also reported that strong leadership at Seattle Children's Hospital has helped bring the different parts of the program together and maintain momentum by showing MCOs that they can collaborate with actuary consultants.

The agreement between Seattle Children's Hospital and the MCOs for a preliminary testing stage has also facilitated the negotiation of the payment model. From April 2016 through the end of the cooperative agreement, the awardee, its actuary consultant, and the four payers have used the preliminary testing stage to refine care management interventions, process measures, and outcome metrics. Actuary staff reported that focusing on the payment methodology framework first has enabled the MCOs to focus on areas of agreement rather than being buried in the details of the methodology. At the end of the preliminary testing stage, Seattle Children's Hospital and the MCOs planned to discuss next steps toward entering into contracts.

Awardee leaders and the actuarial consultants reported three major challenges to developing the payment model: (1) having limited claims data covering the program period, (2) integrating separate data from four MCOs, and (3) developing a model for a relatively small population with a wide range of conditions. First, the awardee noted that due to ramp up in the first year and lags in claims data, it has very little data covering the program period. One awardee leader estimated that calculating program impacts with rigor would require at least one more year of program data. Second, the awardee described how the information systems and analytic resources necessary to receive, normalize, and report on data from four different MCOs have been much more intensive than expected and are a significant challenge for sustainability. However, the awardee leaders view work with multiple MCOs as essential to create a large enough pool of children with medical complexity for a sustainable payment model and program. Third, despite pooling across MCOs, the awardee and actuaries noted that they still faced challenges calculating traditional PBPM and risk adjustment calculations due to the small population with highly variable needs. To address this, they incorporated longer-than-traditional baseline spending assessments (24 to 36 months rather than 12). They and the MCOs also started to discuss with the state Medicaid agency additional options for pooling children in the target population.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The awardee reported plans to only sustain parts of the program it believes have been effective, and did not plan to sustain the intervention as it was originally designed. Seattle Children's Hospital will continue to develop the payment model and engage payers during the no-cost extension period, which ends in May 2018. The awardee did not report plans to scale or replicate its program.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Seattle Children's Hospital had made progress in its goal to engage payers. The awardee had begun testing its payment model with MCOs and made adjustments based on lessons learned from the testing phase. Program staff leaders reported that the program had adapted processes to improve communication and coordination with payers.

C. Implementing the SSR plan: progress and changes

Sustainability. During the third round of site visits, program leaders at Seattle Children's Hospital reported plans to only sustain parts of the program they believe have been effective, and

did not plan to sustain the intervention in its original form. According to the awardee-submitted report from the twelfth program quarter, examples of program activities the awardee found most effective included primary care referrals for care coordination, communication with families regarding patients' complex care needs during transitions of care, and disease-specific interventions (such as the home-visiting program for asthma patients). By the end of the third year, PPIC staff had graduated all of its participants out of the PPIC program, as PPIC intended to minimize families' dependencies on long-term care coordination. Children who will continue to need long-term care coordination services may receive these services from the health home demonstration staff, their PCPs, or payer care managers. Seattle Children's Hospital is also continuing to develop a payment model that might fund other care coordination services for this population.

Scalability. Seattle Children's Hospital did not report plans to scale its program in the third program year, as it was ending the original intervention in the third year.

Replicability. Seattle Children's Hospital did not report plans to replicate its program in the third program year.

D. Factors associated with progress toward implementing the SSR plan

The primary challenge associated with progress toward implementing the SSR plan was insufficient time to prove the program's value, which thwarted the awardee's ability to secure funding and retain program staff. Leaders from Seattle Children's Hospital felt that three years was insufficient to reach agreement with MCOs, because Seattle Children's Hospital could not collect and analyze enough cost and other data to demonstrate effectiveness of the program's services. As discussed in the payment model chapter, the awardee was still negotiating with MCOs around the model at the end of the third program year, and could not definitively sustain program services without funding from MCOs. Without a payment model or other funding in place, program staff began leaving their positions toward the end of the program, according to the awardee. The awardee plans to focus on activities related to implementing its payment model during its no-cost extension period through May 2018.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or SSR plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Seattle Children's Hospital's PPIC program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicaid beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: University of Kansas Health System

April 18, 2018

Colene Byrne (RTI International)
Alison Banger (RTI International)
Brant Morefield (L&M Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The University of Kansas Health System received a 12-month extension to (1) fulfill outstanding contractual obligations, (2) complete an evaluation of the project's impact on quality and cost of care, and (3) complete and present the financial modeling underlying the proposed transitional and transformational payment models for rural health care providers.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

The University of Kansas Health System is using HCIA R2 funds to implement the Kansas Heart and Stroke Collaborative across the state of Kansas. The many Collaborative partners include the University of Kansas Medical Center and School of Medicine, 54 rural hospitals,[4] 12 emergency medical service agencies, a tertiary care hospital, 19 rural physician practices, and two federally qualified health centers (FQHCs). Through one acute care arm and two ambulatory care arms, the Collaborative provides (1) support though clinical protocols, training, and tele-consultation to improve acute care for patients who present to rural hospitals with time-sensitive heart attack, stroke, and sepsis symptoms; (2) short-term (30-day) transitional care management (TCM) for heart attack, stroke, and sepsis patients after discharge; and (3) long-term (up to 11 months) chronic care management (CCM) for patients completing TCM or those who are at high risk for heart attack or stroke who are referred by their primary care physicians or health coaches (Table I.1).

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The University of Kansas Health System received an HCIA R2 cooperative agreement to improve outcomes for heart disease, stroke, and sepsis while reducing the cost of care.
Major innovation	Building trust relationships, program legitimacy, and a long-term strategy for care quality and innovation in rural care delivery with a focus on time-sensitive, high-impact diagnoses
Program components	The program's three arms—(1) ST-elevation myocardial infarction and stroke protocols (also known as the acute care phase of the program), (2) TCM, and (3) CCM—are implemented through the primary components of care management, integrated care, medical home, transitional care coordination, telemedicine, evidence-based clinical practice guidelines, home care, and education and training, as well as the secondary components of patient and family engagement and health information technology.
Target population	Residents of 14 rural northwest KS counties who were hospitalized with or who had symptoms of heart attack or stroke were initially targeted. Residents who were hospitalized with sepsis, at risk for heart attack or stroke, or diagnosed with hypertension or hyperlipidemia were later included in the program. The target population was expanded in Year 2 to cover all of KS. The target population for the acute care phase included all payers. The other two phases targeted Medicare and Medicaid beneficiaries and dual eligibles.
Theory of change/ theory of action	The awardee hypothesized that evidence-based protocols, provider education, telemedicine, TCM and CCM through health coaching, and patient and family engagement would collectively (1) produce measurable improvements in rural Kansans' heart health and post-stroke survival and (2) drive significant reductions in total cost of care related to heart disease, stroke, and sepsis.
Payment model	New fee-for-service payments and shared savings under the Medicare Shared Savings Program. U KS is also designing and analyzing data for CMS on a "transformational" rural health payment model with global budgeting, but the awardee is not implementing the model under the cooperative agreement.
Award amount	$12,523,441
Effective launch date[a]	March 1, 2015

[4] The Collaborative started with 11 critical access hospitals (CAHs) (Cohort 1) in northwest Kansas but has since expanded beyond this area and has recruited additional CAHs and other rural hospitals.

Table I.1 *(continued)*

Program characteristic	Description
Program setting	Rural hospitals, including critical access hospitals; primary care providers; community health care clinics; a tertiary care hospital; an academic medical center; and patients' homes
Market area	Rural
Market location	KS
Target outcomes	• Rate of heart attack and stroke in target population (20% reduction estimated) • Thirty-day acute myocardial infarction (heart attack) mortality rate • Standard heart attack and stroke clinical metrics, such as those endorsed by the National Quality Forum (for example, time to tests, administration of therapy) • Hospital all-cause unplanned readmission, inpatient days after readmission • Medication adherence • Target population discharged alive for heart attack, coronary artery bypass graft, or percutaneous coronary intervention during year prior to measurement year • Rates of ED visits • Rates of transfers to other settings

[a]After the initial planning period, the awardee's program became operational as of this date.

CCM = chronic care management; ED = emergency department; TCM = transitional care management.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the University of Kansas Health System and the Collaborative were successful in implementing their program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors:

1. The awardee enrolled and served 7,334 participants—99.6 percent of its target projection by the end of the three-year cooperative agreement.
2. The awardee delivered all planned program services; closely monitored their implementation; and developed, adapted, and closely followed protocols and policies. The University of Kansas Health System delivered CCM services much earlier than planned and effectively expanded the program across the state of Kansas. By the end of the third year, the Collaborative included 93 health care organizations, which were located in 51 of the 105 Kansas counties.
3. Staffing and training goals were exceeded.
4. The program created an environment of trust and was guided by models of provider engagement. Providers and other participants successfully worked together to improve the continuum of care for the target populations.
5. Health coaches facilitated patient engagement by serving as advocates, educators, and health care navigators through face-to-face, telephone, and tele-health visits. Participating clinicians and non-clinician staff reported that the program had successfully engaged patients and had a positive effect on care delivery.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of the University of Kansas Health System's program, the analysis is still in progress and not

included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. The University of Kansas Health System has initiated participation in a Medicare Shared Savings Program (MSSP) accountable care organization (ACO) to capture savings to sustain the program after HCIA R2 funding ends and to motivate providers to participate in performance improvement initiatives. The awardee has also completed analyses toward development of a transformational payment model that would modify rural hospitals into community care centers with global budgeting and health care resource management on a regional level.

Sustainability plans. The University of Kansas Health System developed a scalable and sustainable program that continues to expand across Kansas. In Year 2, the University of Kansas senior leadership team, in its role as the Collaborative's executive sponsor, approved a long-term sustainability plan. Since the program's inception, the University of Kansas Health System has adopted a long-term strategy and vision for sustaining the program. Strategies were developed not only for program services, but also to support a broader vision of building an integrated care network. Building relationships and establishing trust created the foundation for innovation in the network.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The survey was administered only to clinicians and non-clinician staff who worked with the Collaborative and cared for patients who presented at rural hospitals with heart attack and stroke symptoms. The non-clinician staff survey, which was fielded around the start of the third program year with a sample of 195 potential respondents, achieved a response rate of 85 percent. The clinician survey, which was fielded in the second half of the third program year with a sample of 38 potential respondents, achieved a response rate of 89 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a] Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the University of Kansas Health System was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

Patients were passively enrolled in the acute care intervention when they presented at a rural hospital and received treatment according to protocols for heart attack, stroke, or sepsis symptoms. Some of these patients were covered by payers other than Medicare and Medicaid that are not the focus of this evaluation. Upon discharge after a heart attack, stroke, or sepsis, patients were offered the opportunity to enroll in 30 days of TCM services. The Collaborative actively enrolled patients in TCM as well as CCM services through a form in which they consented to the services. Patients were identified and offered CCM services if they (1) had just completed TCM services, (2) were determined to be at high risk of heart attack or stroke and were referred directly by the health coach or by their primary care providers, or (3) were identified as being at high risk for heart attack or stroke through claims or electronic medical

record (EMR) data. The patients enrolled in TCM and CCM services were Medicare and Medicare-Medicaid dual-eligible patients only. The Collaborative's successful enrollment allowed a long enough period of exposure to the intervention services for any impacts to be detected.

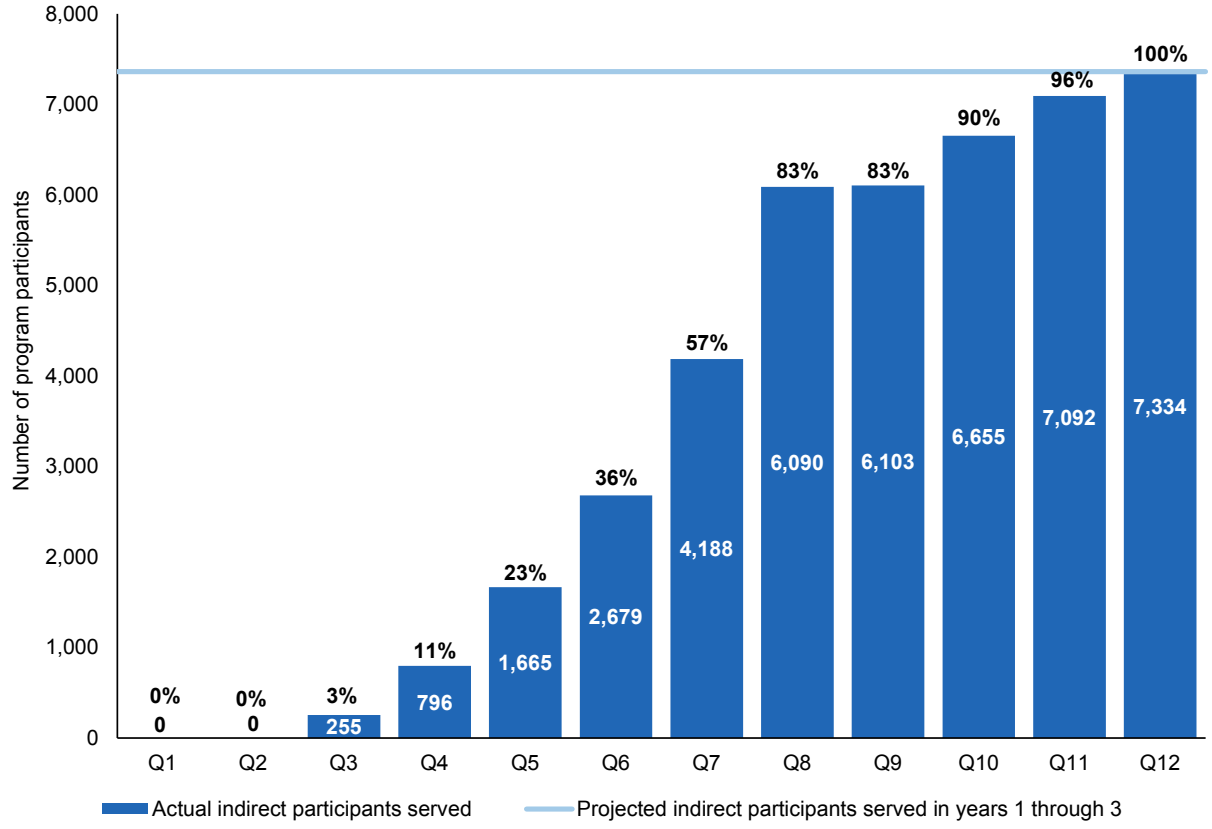
The Collaborative did not make any changes to the way it enrolled patients over the course of the three-year cooperative agreement that would have affected the impact evaluation or interpretation of results. However, some notable adaptations did occur:

- To address difficulty in recruiting patients for TCM services (about half were declining the services), the Collaborative improved its initial communications with patients and family members while patients were still in the hospital. Nurses also visited recently discharged patients who were eligible for TCM to introduce the program to them. In addition, the Collaborative reported that it was more clearly explaining the value of the TCM services to physicians so that they would be more likely to recommend the services for their patients.
- CCM patients could be recruited after their TCM services ended or by being identified by a health coach as high risk and eligible for services. Recruitment was also increasingly through patients' physicians. The Collaborative developed a script for physicians to use to recruit patients for health coaching. A process was added to follow up with the patient by using a letter on the physician's letterhead, so the program was better integrated with the physician's practice.
- Health information technology (IT) implemented at the end of the second year improved enrollment by creating a registry of high-risk individuals to offer CCM services to.

b. Evidence of enrollment effectiveness

The Collaborative was effective in meeting its enrollment target by the end of the 12th program quarter. Overall, the awardee reported that it served 7,334 indirect participants from March 2015 (when it launched its program) through August 2017, which represents about 99.6 percent of its final three-year projections (Figure II.1). Improved identification of potential CCM patients and direct engagement with primary care physicians who referred TCM and CCM patients were key to the awardee's successful enrollment.

Figure II.1. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers.

c. Barriers and facilitators associated with enrollment effectiveness

The Collaborative's progress in meeting its three-year enrollment goals was influenced by several barriers and facilitators. Barriers to enrollment identified by program leaders included difficulty working with providers to actively encourage referrals of potential patients and provider mistrust about the legitimacy of the program. Although the Collaborative was hindered in identifying patients for CCM services early in the second year due to lack of claims data, other data sources were used and there were no delays in implementation. In Year 2, new population health IT tools for accessing EMR data greatly improved the Collaborative's ability to identify high-risk individuals. The CCM services were implemented in the second program year, a year earlier than planned.

The Collaborative successfully conducted outreach with potential participants. Adding recruitment sites such as primary care clinics facilitated enrollment. The number of "warm handoffs" between the providers and health coaches increased and helped with recruitment. Early in the program, the Collaborative developed and refined the health coach and transitional

care manager roles, which led to more effective recruitment. The Collaborative expanded the scope of its acute care protocols (for example, by adding sepsis and trauma) as well as hospital, physician, and other provider participation across the state, which expanded the number of potential participants.

2. Delivery of program services

a. Description of and changes to service delivery model

The Collaborative sought to improve outcomes and reduce the cost of care for Kansans who have a diagnosis of, symptoms of, or are at risk for heart disease, stroke, or sepsis through three new service delivery arms—an acute care arm and two ambulatory care arms (TCM and CCM).

The Collaborative was highly successful in the delivery of program services across all three arms. Overall, the program design and the limited changes made to it allowed the University of Kansas Health System to quickly scale up its implementation effort. Although the Collaborative expanded the number of conditions it addressed and made some program enhancements as it developed best practices, the delivery of program services across the three arms and across implementing sites was consistent and effective. This consistency supported effective replication in new provider sites and communities.

In the acute care arm, patients of rural hospital EDs received services according to acute care protocols that standardized care in the event of suspected heart attack, stroke, or sepsis (which was added in the second year). In addition, clinicians received help in treating these conditions from remote emergency or critical care specialists by using Avera's eCare[5] eEmergency tele-health solution. The main changes to the acute care arm over the three years were (1) adding sepsis in the second program year as another time-sensitive condition to treat according to protocols, (2) adding acute care heart failure clinical guidelines in the middle of the third program year, and (3) implementing evidence-based protocols for trauma and patients at the end of life in the last quarter of the third program year.

At the direction of its Clinical Committee, the Collaborative also facilitated a statewide, multiple stakeholder collaboration to develop, pilot, and refine a protocol to standardize and improve the reliability and quality of the information exchanged between rural providers who transfer patients and the hospitals and specialists that receive the patients.

Starting at the end of the first year, Collaborative staff traveled every quarter to all participating rural hospitals to conduct data collection. This data collection approach allowed staff to understand each site's medical record systems and nuances, gain important background and contextual insights, and provide immediate feedback to participating sites about their performance. These rural providers benefitted from having Collaborative staff extract their

[5] Avera eCare Services is a suite of distance telemedicine equipment and services affiliated with Avera Health, a large health system based in Sioux Falls, South Dakota. (See http://www.averaecare.org/ecare/.) The Collaborative is using Avera eCare's eEmergency services, in which board-certified ED physicians and critical care nurses deliver immediate, supportive care and nursing documentation to CAH EDs. (See https://www.averaecare.org/ecare/what-we-do/emergency/.)

medical record data and provide feedback reports that were designed to enable data-driven improvements.

Reviewing medical records and other data with participating hospitals was an important part of the performance improvement process. In the 10th program quarter, the Collaborative became a patient safety organization (PSO) listed by the Agency for Healthcare Research and Quality.[6] Nearly all of the participating hospitals became part of the PSO. The Collaborative worked with participating providers in interorganizational performance improvement activities under the privilege and confidentiality protections of the Patient Safety Act of 2005 and relevant regulations.

The other two arms are ambulatory care services. The Collaborative provided TCM services to patients for 30 days following discharge from the hospital and CCM services to patients who were (1) transitioning out of TCM services, (2) primary care outpatients at risk for heart attack or stroke who had been referred directly by their primary care providers, or (3) identified as at risk by the Collaborative's population health IT tools.

During the 30-day TCM period, the transitional care managers, who are advanced practice nurses, completed a home visit and made follow-up telephone calls (usually six to eight calls) with patients to monitor their health, ensure that they understood and followed their medication regimen, and facilitate follow-up visits with their primary care physician and other specialists. The transitional care managers worked with their local hospital's discharge planner to ensure that patients had a comprehensive and integrated plan when they were discharged from the hospital.

CCM services, delivered by a health coach (usually a registered nurse), included a home visit and telephone calls to assess the physical and psychosocial well-being of participants. As noted earlier, patient referrals to CCM from physicians increased over the program. The Collaborative developed a script for physicians and other processes so that the CCM services were integrated into the physician's practice.

As part of the TCM and CCM services, the Collaborative implemented evidence-based ambulatory care protocols in the second program year for treatment of patients with hypertension, hyperlipidemia, and tobacco use.

TCM and CCM patients who were recovering from a heart attack or heart failure were provided with a formal resiliency training program by trained health coaches, in order to help them and their family members better understand the disease. They also taught patients strategies for managing symptoms and skills to cope with functional limitations.

Early in the project, the Collaborative developed a spreadsheet program for CCM tracking. This allowed the Collaborative to get the CCM services up and running, as well as to define the

[6] Patient safety organizations (PSOs), as defined under the Patient Safety Quality Act of 2005, are entities that do not have public or private regulatory responsibilities; are able to collect and analyze data; have mechanisms for communicating with health care facilities, providers, and plans; are linked with initiatives for conducting research and demonstrations to improve health care quality; and, most importantly, have legal protections for the identities of the providers involved in a patient safety event or its reporting. Entities wishing to become a PSO must apply through the Agency for Healthcare Research and Quality for certification. (See https://www.pso.ahrq.gov.)

data and task tracking needs. As the CCM services were expanded, the Collaborative identified, implemented, and customized commercial software to better support these functions. All the health coaches were trained in and used the care manager software. This second-generation care management software helped accelerate program expansion and increased the health coaches' productivity.

At the end of the third year, the Collaborative fully implemented "eHealth coaching" by using a team of trained coaches to provide CCM services using tele-health technology. These eHealth coaches served communities in which the Collaborative was unable to recruit or maintain a health coach, thereby expanding the reach of the program.

Although the use of health IT tools was originally planned for the program, the creation of registries and population health management reports specifically was considered well beyond the capabilities of individual sites. Population health management expanded after the Collaborative secured a vendor[7] in the second program year to support claims analysis, EMR extraction, and clinic performance profiling. The Collaborative developed patient registries for specified conditions that elevated the risk of heart attack or stroke, including post–heart attack and post-stroke diagnoses, hypertension, hyperlipidemia, and tobacco use. These registries allowed the Collaborative to monitor the percentage of patients who were treated according to guidelines compared to the goal (for example, the goal levels for blood pressure or cholesterol treatment) by leveraging technology and data analytics to drive performance improvement. A population health dashboard supported analysis at the regional, site, and individual patient levels. The dashboard incorporated data from over 15 different EMR systems.

In addition, the Collaborative provided, by using a variety of means and staff, ongoing education to improve community awareness of the signs and symptoms of heart attack, stroke, and sepsis.

The Collaborative's services were highly successful in facilitating the integration of rural health services by providing clinical programs for heart attack, stroke, sepsis, heart failure, trauma, and palliative care across the care continuum, with the goal of improving outcomes and reducing the cost of care for Kansans with the targeted conditions.

According to the theory of change/theory of action (TOC/TOA) for the acute care arm, patients who presented with time-sensitive symptoms of heart attack, stroke, or sepsis were treated by emergency medical services (EMS) and hospital staff who recognized these symptoms and who were trained in and used clinical guidelines-based protocols. These protocols were expected to result in more rapid transport to hospitals, faster administration of evidence-based diagnostic tests and therapies, reduced mortality and morbidity from these conditions, and reduced health care costs. Better care coordination and patient education by the acute care staff helped patients transition from the hospital and improved their understanding and management of their conditions and medications. Based on interviews, staff and clinician surveys, and self-monitoring measures, the Collaborative appears to have met all of the important elements of the TOC/TOA to achieve its target outcomes for the acute care arm. The Collaborative described, for

[7] The Collaborative selected Lewis and Clark Health Information Exchange and its partner, Health Metrics Systems.

example, cases of patients who received stroke therapies in a timely manner that saved lives and prevented the long-term effects of a stroke.

In the TOC/TOA for the TCM and CCM arms, patients who had a recent heart attack, stroke, or sepsis or who were considered at high risk of a heart attack or stroke received care coordination and health coaching. Patients who received coaching were expected to have improved health literacy (including understanding their conditions and medications), engagement, and resiliency, as well as better self-management of their conditions. These improvements were expected to reduce morbidity and mortality from these conditions and the associated health care costs (for example, hospital admissions, readmissions, and ED visits). Based on interviews, staff and clinician surveys, and self-monitoring measures, the Collaborative met all of the important elements of the TOC/TOA to achieve the target outcomes for the TCM and CCM ambulatory care arms.

b. Evidence of service delivery effectiveness

Overall, the Collaborative was successful in achieving service delivery effectiveness across all three arms. The Collaborative hired and trained staff and began to provide services under the acute care and TCM arms early in the first program year. The Collaborative then accelerated CCM services and delivered them early in the second year. The program was successfully expanded in rural communities across the state and the number of participating providers grew quickly. Positive word of mouth from early adopters made direct recruitment largely superfluous.

Service delivery was effective largely because the program was conditioned upon (1) getting the right players together; (2) buy-in, trust, and collaboration; (3) understanding the resource constraints of rural providers; and (4) adaptability. HCIA R2 funding allowed the creation of a "purposeful infrastructure" and gave legitimacy to the program. The Collaborative anticipated that staff from participating hospitals and other providers would have limited time to dedicate to project activities. If the project were to be successful, it would have to work within this constraint. Adaptability was key to understanding what was effective and what was not and to making important program changes along the way.

Delivery of intervention services. The awardee delivered the services as planned but made a number of refinements along the way, as described earlier. Ninety-seven percent of clinicians surveyed felt that the heart attack and stroke protocols arm of the program, which was the focus of the survey, had a positive impact on the quality of care and services they provided to patients. Similarly, 93 percent of staff surveyed indicated that the program had a positive impact on the quality of care and services provided to participants. Half of the staff surveyed strongly agreed and another 38 percent somewhat agreed that the program was making a difference in meeting critical needs in the community.

Staffing and training. The Collaborative was successful in recruiting, hiring, and retaining the program staff needed to deliver services. Providing training, supervision, and support for program staff and fostering communication and collaboration among staff were identified by leaders as major facilitators to the delivery of program services, while difficulty hiring or retaining staff was reported to be a minor barrier. Similarly, only 17 percent of surveyed staff and 13 percent of surveyed clinicians reported that hospital staff turnover or unfilled positions posed a major barrier to achieving program goals.

The University of Kansas Health System reported that, as of August 2017, it had exceeded its hiring goal and had achieved a 93 percent retention rate for staff. In addition, the Collaborative had trained 1,136 trainees (including non-clinician staff and clinicians) since the project's inception—far exceeding its three-year training target.

Seventy-two percent of staff surveyed early in the third program year reported that they had attended formal training and 84 percent reported that they had received informal instruction related to the program. Informal training was provided through staff meetings (90 percent); self-study (64 percent); asking colleagues for help (74 percent); mentoring, shadowing, or technical assistance (around 40 percent each); and individual (42 percent) and group supervision (48 percent). Most of the surveyed non-clinician staff strongly agreed (47 percent) or somewhat agreed (36 percent) that the training helped them improve their job performance.

All of the clinicians surveyed early in the third program year reported that they had attended formal training. Most reported that they had also received informal instruction related to the program. Clinicians reported that their formal training included the goals of the program (88 percent), chronic care management (75 percent), quality of care (94 percent), team-based care (94 percent), and performance improvement (78 percent). Informal training was provided through a variety of means including staff meetings (90 percent), self-study (61 percent), asking colleagues for help (71 percent), meeting with program administrators (69 percent), and group supervision (38 percent). Most of the surveyed clinicians strongly agreed (53 percent) or somewhat agreed (28 percent) that the training helped them improve their job performance.

Recruitment and engagement of providers. The Collaborative was very successful in recruiting and engaging providers across Kansas. The Collaborative had a wide range of partners that participated in various facets of the program. The University of Kansas Health System and the University of Kansas Medical Center Schools of Medicine, Nursing, and Health Professions provided clinical support for the development of the acute care protocols and support for training. Hays Medical Center (HaysMed)—a rural tertiary care hospital in Hays, Kansas—served as a local convening organization. Eleven critical access hospitals (CAHs) in northwest Kansas served as the original implementing cohort sites for the acute care protocols. The Collaborative expanded significantly over the cooperative agreement, with a sevenfold increase in participants since the first year. As of August 2017, there were 46 rural hospitals, 19 physician practices (including rural health clinics), 21 EMS and air medical transport providers, and two FQHCs participating in the Collaborative or the Kansas Clinical Improvement Collaborative (KCIC). In addition, 29 provider organizations in Kansas and one in Missouri expressed interest in or received a formal presentation regarding participation in the Collaborative.

The awardee retained all of the Collaborative participants for continued participation after the award either through the ACO or PSO that was spun off from the Collaborative.

Based on interviews, hospitals and providers liked the direct connection to evidence-based practices and appreciated how the Collaborative linked them to protocols that helped them use best practices and offer a high level of care. In the past, it may have taken years

> "They made it easy for us to implement the program. They brought the training and education to us. They come out and assist with the quarterly data gathering. This [program] really is as advertised."
>
> —Participating rural hospital official

to get new guidelines implemented as standards of care. Providers liked the knowledge transfer and staying engaged with the academic medical center.

Primary care clinicians became increasingly engaged in and referred their patients to the TCM and CCM programs. A physician champion who works out of a CAH reported being thrilled with the work the health coach does with his patients, noting that the coach has gotten numerous resources for his patients and made a positive difference in their lives. This physician also reviewed his hospital's feedback reports on sepsis. He and other hospital staff learned that they were not performing well in following the sepsis protocols, in part because they were not diagnosing and documenting sepsis. They worked on improving recognition, documentation, and treatment of sepsis.

Engagement of program participants. The Collaborative engaged program participants in several ways. Raising awareness about the health coaching program was important to patient acceptance. The coaches used consistent terminology and clearly identified the TCM and CCM programs, which made it easier for patients to understand that this was a cohesive set of services, which they were more willing to accept. The resiliency program provided to patients who had suffered heart failure and their families was well received.

According to interviews and survey results, the awardee successfully engaged patient participants in the TCM and CCM arms. Seventy-one percent of surveyed staff agreed that the program successfully engaged participants. Patients surveyed by the program at hospital discharge reported being prepared for their care transition. In the 12th program quarter, 68 percent of patients strongly agreed that staff took their preferences and those of their family or caregiver into account in deciding what their health care needs would be when they left the hospital. Sixty-four percent of patients strongly agreed that they had a good understanding of the things they were responsible for in managing their health and 69 percent strongly agreed that they clearly understood the purpose of each of their medications when they left the hospital.

c. Barriers and facilitators associated with service delivery effectiveness

The Collaborative's ability to effectively deliver intervention services was influenced by several facilitators and barriers, as discussed below. Facilitators included strong leadership, program support for participants, understanding of rural health and the need to build trusting relationships, minimizing provider burden to participate, adaptability, performance monitoring and feedback, population health tools, and protocols for the standardized transfer of information. Barriers including health coach staffing and retention and initial clinician buy-in as discussed below.

There was strong support from program leaders, partners, and officials at the University of Kansas Health System, who brought to the program a deep understanding of rural health and the need to find ways to deliver evidence-based care by using resources available to rural providers. Clinical guidelines and practices were adapted to rural resources and circumstances. A related strong facilitator was that the Collaborative established an environment of trust and recognized trusting relationships as the foundation for the program and for innovation.

The respondents interviewed attributed much of the Collaborative's successes to the strength and commitment of its staff and their connection with rural health in Kansas. Program leaders

worked extensively with sites to help them understand the Collaborative's goals and how they could better participate. The executive director alone traveled many tens of thousands of miles across the state to meet with site leadership, physicians, and others to promote the Collaborative's work and secure their participation. Although not a majority, a substantial proportion of staff surveyed (42 percent) felt that this support was one of the most helpful factors for achieving the program's goals.

The program staff provided support and encouragement to the providers while holding them accountable for better outcomes. They recognized and celebrated large and small successes and offered resources when things weren't going as well as hoped—for example, when providers fell back into old habits and were not following the protocols. They also held the local sites responsible for their activities and meeting the performance benchmarks.

> "It's easy for us to come in and say, 'Here's the evidence-based guidelines and here's an order set based on those, now just do it.' It takes the willingness of the local system to implement them and then to be willing to sit down, to listen to the critique on what the data [are] showing, where there's opportunities for performance improvement, to embrace that need for change and not just fall back to 'but this is the way we've always done it.' It's the local participating sites that have made this successful because they're the ones doing the work at the frontlines."
>
> —Collaborative executive director

As described earlier, the population health tools that the Collaborative deployed to support claims analysis, EMR extraction, patient registries, and clinic performance profiling facilitated service delivery by identifying potential TCM and CCM patients and supporting performance improvement. The executive director of the Collaborative worked with a vendor to create meaningful dashboards on the care of post–heart attack and post-stroke participants and participants with hypertension, hyperlipidemia, and tobacco use. The dashboards were shared with the clinics and showed them how well they were following evidence-based practice.

Adaptability was a key part of the Collaborative's strategy for success. Collaborative staff continuously measured and monitored the program's effectiveness and received feedback from participating providers. For example, as the Collaborative expanded its reach across Kansas, the service delivery model was refined to streamline the training process for new sites and simplify procedures for enrolling patients in CCM.

When performance worsened on four of the six hospital process measures in the 10th program quarter, the Collaborative stepped in to reinforce the initial work that was done with these hospitals. The quality improvement staff visited each of the 43 participating hospitals during the next quarter. This experience reinforced the need to have one-on-one meetings with all hospitals every quarter to review the hospital's performance and address any questions or concerns. A Collaborative staff member helped the hospital team thoroughly analyze outlier cases to identify opportunities for improvement and develop appropriate action plans.

> "Part of our success can be attributed to really taking a look at what the evidence base tells us is key to success and constructing a program that wasn't necessarily how we envisioned it in the beginning but evolved based upon those findings. Just with trial and error, we went this way, two steps forward, and found that's not quite working. Let's take two steps sideways and keep moving. It's that adaptability that makes us successful."
>
> —Program manager

Another example of adaptability was the Collaborative's creation of the PSO when staff and providers identified the need to improve hospital quality by reviewing patient care and records in a protected environment with more in-depth patient safety reviews, including root cause analyses in the case of medical errors. Without the protections afforded by a PSO, sites were reluctant to complete these analyses with the Collaborative because of liability concerns.

When rural hospitals raised concerns about the reliability and quality of communications between transferring and receiving facilities, the Collaborative facilitated the development of a standardized transfer communication protocol. Under the protocol, a transferring facility agreed to complete and send with the patient the standard form. The receiving hospital committed to sending to all specified individuals at the transferring hospital a patient progress note within 24 to 48 hours of the patient's arrival and a patient summary of care within 24 to 48 hours of discharge.

Although the Collaborative was successful in hiring and retaining trained staff, retaining nurses for TCM was a challenge because of the regional nature of the work and the amount of time that the transitional care managers had to spend on the road. Rural communities generally have a difficult time hiring and retaining clinical staff. In response, the Collaborative provided support for TCM services for local clinics that wanted to provide TCM for their patients. The support included training, access to the Collaborative's processes and forms, and assistance with any regulatory or accreditation compliance issues. The Collaborative also expanded its reach in communities where it had been unable to recruit or maintain a health coach by providing CCM services through tele-health technology.

One of the facilitators to staffing was hiring a local program manager, who helped maintain connections between the Collaborative and the broader network of resources at the University of Kansas's academic medical center (where experts in stroke care, sepsis, and so on help develop, review, and refine the acute care protocols). She was a significant contributor to the team while the executive director and operations manager spent more time traveling to local communities.

There was variation in physicians' willingness to engage and refer patients. Although some were quick to see the benefits for their patients, others were more reluctant to engage with the program. The Collaborative refined its process for reaching potential patients by placing minimal burden on busy physician practices. Clinic staff spent less than five minutes per beneficiary in outreach to potential patients.

There were also challenges in getting some rural hospital physicians to buy into the program and use the acute care protocols. Coordinating with the different cardiology practices around the state (and in neighboring states) was challenging because cardiologists often have slightly different approaches to how they provide care. The Collaborative's executive director and medical director spent time with the cardiologists to review the protocols, explain the best practices, and gain buy-in for the use of the protocols.

Buy-in of the acute care protocols by hospital clinicians (for example, nurse practitioners and physicians) was identified by 78 percent of surveyed clinicians as one of the most helpful factors in improving heart attack and stroke care. A smaller percentage of surveyed non-clinician staff, 38 percent, felt that clinician buy-in was one of the most helpful factors. Only a very small

percentage of surveyed non-clinician staff and clinicians felt that resistance to the acute care protocols, either by non-clinician hospital staff or by clinicians, was a major barrier to improving heart attack and stroke care. Resistance was identified as a minor barrier by almost 25 percent of surveyed staff and clinicians, suggesting that there was some early resistance, but it was not great. In addition, 94 percent of surveyed clinicians felt that the workload for using the acute care protocols was about right; none reported that it was too or somewhat heavy. Seventy-one percent of clinicians were very satisfied and another 23 percent were somewhat satisfied with their role in using the acute care protocols. Further, 75 percent of clinicians felt they were an important member of the team that used the protocols and 72 percent strongly agreed that they were supported by their colleagues and management to use the protocols.

C. Assessment of perceived program effects on the delivery of care and outcomes

The majority of rural hospital non-clinician staff who responded to the survey at the beginning of the third program year believed that the acute care arm of the Collaborative program had positive impacts on care delivery and outcomes such as patient satisfaction and quality of life (Table II.2). Most staff surveyed either strongly (50 percent) or somewhat (38 percent) agreed that the program was making a difference in meeting critical needs in the community. In addition, most staff indicated that the program was very (54 percent) or somewhat (32 percent) effective in achieving its goals. Eighty-seven percent felt the program was worth the effort.

Table II.2. Staff perceptions of Collaborative acute care arm effects on care

The Collaborative acute care arm had a positive impact on:	Percentage
The quality of care and services you provide to participants	93
Your ability to respond in a timely way to participant needs	88
Your ability to provide care or services that are responsive to participant preferences, needs, and values	90
Access to care or services for all participants	80
Achievement of participants' health goals	88
Participant satisfaction	84
Participant quality of life	89
Care coordination	86

Source: HCIA R2 evaluation survey of non-clinician rural hospital staff involved in implementing the acute care protocols, fall 2016

Note: The survey had 148 respondents overall.

Most clinicians who responded to the clinician survey during the second half of the third program year believed that the Collaborative program had positive impacts on the delivery of care and outcomes such as patient satisfaction and quality of life (Table II.3). Most clinicians surveyed either strongly (66 percent) or somewhat (19 percent) agreed that the program was making a difference in meeting critical needs in the community. In addition, the clinicians indicated that the program was either very (81 percent) or somewhat (19 percent) effective in achieving its goals. All (100 percent) felt the program was worth the effort.

Table II.3. Clinician perceptions of the Collaborative's effects on care

The Collaborative had a positive impact on:	Percentage
The quality of care and services you provide to participants	97
Your ability to respond in a timely way to participant needs	100
Your ability to provide care or services that are responsive to participant preferences, needs, and values	94
Access to care or services for all participants	84
Achievement of participants' health goals	94
Participant satisfaction	91
Participant quality of life	100
Care coordination	97

Source: HCIA R2 evaluation of rural hospital clinicians involved in implementing the acute care protocols, winter 2017.

Note: The survey had 32 respondents overall.

The awardee's own self-monitoring measures and analyses provide another picture of the program's effectiveness. The Collaborative monitored rural hospital performance on six key performance metrics that measured time from arrival to intervention for heart attack and stroke symptoms. Results stabilized over the last several quarters, with response times significantly below their pre-intervention baselines. Hospitals have also shown improvement on measures of compliance with the clinical guidelines for ST-elevation myocardial infarction and stroke. The Collaborative notes that "time is muscle" with heart attack, "time lost is brain lost" with stroke, and improved performance translates into lives saved and quality of life maintained.

The awardee's analyses of claims data suggested that there may have been some early cost savings related to the program. The University of Kansas Health System built a custom-designed database of Medicare Parts A and B claims and compared the costs of care in 2013 (before program implementation) with the costs of care in 2016 (after implementation) for beneficiaries in the initial participating counties in northwest Kansas with the original 11 Cohort 1 CAHs, to show potential cost savings related to the interventions. The analyses included a comparison of the costs of an acute myocardial infarction episode in 2013 to the costs in 2016. The costs were broken down into inpatient, outpatient, and professional categories and by disease and risk categories, such as whether the patients had heart failure, heart attack, hypertension, or were smokers. The awardee found statistically significant reductions in both the number of heart attacks and the total cost of care (2014–2016 combined) for those who suffered heart attacks in the 11 counties where the Cohort 1 CAHs were located.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

The impact evaluation will estimate the impact of the program on Medicare beneficiaries in the Cohort 1 counties in rural northwest Kansas. The analyses will focus on beneficiaries who were hospitalized with heart attack or stroke. These patients may have been treated according to the protocols. In addition, some of these patients may also have received TCM or CCM services.

The treatment and comparison groups capture only Medicare beneficiaries, who make up most of the patients who are treated at the rural hospitals.

The Collaborative was very successful in meeting its goals and in delivering services across the state, including in rural northwest Kansas. Program leaders and staff reported that the awardee enrolled participants early enough in the program to receive a sufficient amount and duration of services to expect improvements in clinical care. These improvements are predicated on the TOA/TOC being accurate and reflecting planned processes that can result in the desired outcomes.

Evidence from the Collaborative's performance process measures indicates that the acute care protocols were implemented and followed by the Cohort 1 CAHs. Six of the measures have shown significant improvement from baseline, while a few have shown little improvement or some slippage.

The main effects of this program are likely to be a reduction in ED visits, all-cause hospitalizations, and 30-day unplanned readmissions, leading to a decrease in total cost of care. Based on the TOA/TOC, there would also be an expected reduction in mortality, to be measured as the decrease in the probability of death within 90 days of admission.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of the University of Kansas Health System's program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: University of Kansas Health System

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	1,833[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,649
Likelihood of all-cause hospitalizations	1,423
MDE sample size requirement to detect 20% effect	
Total expenditures	412
Likelihood of all-cause hospitalizations	356
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	No serious issues; proceeding with comparison group selection
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician, staff and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We plan to conduct a rigorous impact analysis of the University of Kansas Health System's intervention for heart attack and stroke. Because the initial intervention has been extended to other rural areas in the state, we have selected 23 CAHs in rural counties in Nebraska as a source of comparison beneficiaries. We expect the two groups to be broadly well matched at the outset, but we will compare demographic characteristics and prior use for beneficiaries presenting in Kansas and in Nebraska to ensure that the treatment and comparison groups are comparable.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents a summary of the baseline characteristics of the treatment group, measured during the 12 months before the enrollment date for each beneficiary, which is the date of receiving acute care services for stroke or heart attack at a facility associated with the awardee. More precisely, for the purpose of our evaluation, the treatment group consists of individuals who present at rural hospital EDs—in both the CAHs and the supporting tertiary hospital for the respective health network, HaysMed—with symptoms of a stroke or heart attack.

The original facilities participating in the University of Kansas Health System program started applying acute care protocols for stroke and for heart attack in March 2015. By the end of May 2016, there were 1,334 Medicare enrollees who had received those acute care protocols and had enough information provided by the awardee to be successfully matched to Medicare claims.[8] Ninety beneficiaries were excluded from the analysis because they were not enrolled in Medicare fee-for-service (FFS).

In presenting baseline characteristics for this report, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, Parts A and B, with Medicare as the primary payer at the time of their University of Kansas enrollment date, who had met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment) and who could be identified in the Virtual Resource Data Center's Medicare health insurance claim-to-beneficiary ID crosswalk. The calendar period covered by the baseline quarters is determined by the enrollment date for each participant and therefore varies by participant. After we excluded beneficiaries who did not meet the above criteria, a total of 1,244 participants were included in the analysis of baseline characteristics for this report.

The results of our analysis on baseline demographic characteristics (Table III.2) indicate that the Medicare FFS population enrolled in the University of Kansas Health System program is predominately white and older (that is, 65 and older). There is one beneficiary enrolled in hospice and very few beneficiaries (n = 4) with end-stage renal disease (ESRD). As expected, the hierarchical condition categories (HCC) scores of program participants are higher than the national average, with a mean of 1.74 and a median of 1.32.

[8] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the University of Kansas's program through May 31, 2016

	All participants (N = 1,244)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	142	11
65 to 74	419	34
75 to 84	432	35
85 and older	251	20
Gender		
Female	571	46
Male	673	54
Race		
White	1,229	99
Black	5	0.4
American Indian, Alaska Native, Asian/Pacific Island American, or other	2	0.16
Hispanic	5	0.4
Original reason for Medicare eligibility		
Old age and survivor's insurance	980	79
Disability insurance benefits	260	21
ESRD[a]	4	0.32
Hospice[b]	1	0.08
Medicare/Medicaid dual status, percent dual[b]	195	16
HCC score[c]		Statistic
Mean		1.74
25th percentile		0.79
Median		1.32
75th percentile		2.26

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which a beneficiary was admitted to a participating hospital and was diagnosed as having a heart attack or stroke. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3 shows baseline expenditure and utilization data for a common set of measures for the four quarters prior to enrollment. The higher-than-average HCC scores are consistent with the high total Medicare expenditures per beneficiary per month (PBPM) shown in the table. On average, the PBPM expenditures during the 12 months before enrollment were $1,382—substantially higher than the national average Medicare expenditure of about $790 per month. Multiple spending categories drove high costs, including inpatient ($433 PBPM), outpatient ($434 PBPM), physician ($227 PBPM), and skilled nursing facility ($171 PBPM) services.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the University of Kansas Health System's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	1,244	1,219	1,229	1,244	1,244
Average Medicare expenditures PBPM[a]					
Total	1,382 (59)	1,109 (72)	1,241 (85)	1,393 (94)	1,775 (108)
Acute inpatient	433 (27)	342 (41)	366 (44)	441 (49)	581 (58)
Inpatient other[b]	24 (7)	17 (10)	38 (16)	16 (10)	24 (11)
Outpatient[c]	434 (19)	398 (31)	424 (24)	406 (21)	509 (24)
Physician services	227 (9)	191 (11)	203 (12)	232 (14)	279 (16)
Home health	40 (5)	29 (5)	45 (7)	40 (6)	46 (6)
Skilled nursing facility	171 (20)	85 (20)	115 (24)	204 (41)	275 (47)
Hospice	1 (1)	0 (0)	0 (0)	0 (0)	3 (3)
Durable medical equipment	53 (4)	47 (4)	50 (4)	55 (5)	59 (5)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	585 (30)	445 (42)	536 (51)	562 (49)	791 (61)
Outpatient ED visits	1,068 (111)	962 (118)	1,073 (150)	995 (123)	1,235 (119)
Observation stays	187 (15)	122 (20)	164 (24)	204 (29)	257 (31)
Primary care visits in any setting	5,119 (208)	4,466 (233)	5,021 (256)	4,929 (267)	6,004 (287)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Primary care visits in ambulatory settings	3,458 (152)	3,124 (164)	3,418 (172)	3,335 (166)	3,907 (184)
Specialist visits in any setting	7,546 (266)	6,822 (292)	7,347 (352)	7,366 (324)	8,612 (352)
Specialist visits in ambulatory settings	6,080 (201)	5,777 (229)	6,025 (261)	5,902 (246)	6,589 (247)
Measures of any health care utilization					
Percentage with a hospital admission[d]	34 (1)	10 (1)	10 (1)	12 (1)	15 (1)
Percentage with an outpatient ED visit[e]	45 (1)	15 (1)	16 (1)	15 (1)	20 (1)
Percentage with an observation stay[f]	15 (1)	3 (< 0.5)	4 (1)	5 (1)	6 (1)
Percentage with a 30-day readmission among all discharges	15 (1)	9 (3)	10 (2)	16 (3)	21 (3)
Percentage of participants with a readmission among all participants	5 (1)	1 (< 0.5)	1 (< 0.5)	1 (< 0.5)	2 (< 0.5)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

Fourth quarter expenditures, spanning the three months before enrollment, were notably higher than for earlier quarters ($1,775 PBPM in quarter 4). In this quarter closest to enrollment, average expenditures were highest across most expenditure categories—notably so in acute inpatient ($581 PBPM and 34 percent higher than the 12-month average), outpatient ($509 PBPM and 17 percent higher than the 12-month average), physician ($279 PBPM and 23 percent higher than the 12-month average), and skilled nursing facility ($275 PBPM and 61 percent higher than the 12-month average) services.

Increases in measures of utilization accompanied the rise in fourth quarter expenditures. The average rates per 1,000 beneficiaries over all 12 baseline months included 585 acute hospital admissions; 1,068 outpatient ED visits; 187 observation stays; 5,119 primary care visits; and 7,546 specialist visits. The rates of utilization were highest for the enrolled population in the fourth quarter, closest to enrollment, over all utilization types. Of particular note, rates of acute hospitalization, outpatient ED visits, and observation stays in the fourth quarter were 35 percent, 16 percent, and 37 percent higher, respectively, than the 12-month averages. The increases in these utilization measures suggest that the enrolled heart attack and stroke patients were incurring hospital-related charges just prior to the enrollment event.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

The University of Kansas Health System has initiated participation in a MSSP ACO to capture savings to sustain the program after HCIA R2 funding ends and to motivate providers to participate in performance improvement initiatives. The awardee has also completed analyses toward development of a transformational payment model that would modify rural hospitals into community care centers with global budgeting and health care resource management on a regional level.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

The University of Kansas Health System implemented two payment approaches to support services after the end of the cooperative agreement: (1) billing of Medicare FFS TCM and CCM codes and (2) participating in an MSSP ACO. First, the awardee expected to fund the health coaches and maintain the health IT infrastructure through billing of TCM and CCM codes. Second, the awardee implemented an MSSP ACO to capture and use shared savings to sustain the program and motivate providers to continue participating in performance improvement initiatives (for example, participation in the PSO to improve patient care and safety or dashboard reports of performance on MSSP quality measures).

With approval from CMS, the University of Kansas Health System conducted analyses for further design and development of its proposed transformational rural health care multi-payer model, which would involve global budgeting and health care resource management on a regional level. The model was intended to fund and rightsize health care services in rural communities in a more flexible manner than allowed by current payment rules while at the same time delivering value to rural beneficiaries and cost savings to CMS. It would involve regional planning with multiple facilities, each serving their immediate area. Specifically, the Community Care Center model would involve Medicare, Medicaid, and commercial beneficiaries in the immediate area surrounding a facility (for example, a CAH or rural community hospital). Under the model, financially distressed hospitals would convert to outpatient providers with 12- or 24-hour emergency services and a small number of transitional or "swing" beds. They would offer specific services (physical and respiratory therapy, diagnostic imaging, and clinical laboratory services) and engage in population health improvement activities. The awardee believed this model would enable rural communities to maintain local access to essential health care services while realizing savings for the Medicare and Medicaid programs. Citing research conducted by the University of North Carolina's Cecil B. Sheps Center for Health Services Research, iVantage Health Analytics, and others, the awardee said that cost savings would result from lower inpatient costs after rural hospitals' conversion to outpatient facilities as well as from tailoring the services available based on identified community need. According to the awardee, nearly one in three rural hospitals in Kansas faces closure due to financial distress. Converting these

distressed facilities could avoid the increases in costs associated with lack of access to local care, which often occur after failing CAHs are closed.

C. Status of the payment model

The awardee planned to continue with the billing of Medicare FFS TCM and CCM codes. The awardee also planned to expand providers' participation in the MSSP payment model after the cooperative agreement and its no-cost extension end through a new legal entity, the KCIC, which was created to meet the MSSP regulatory requirements and to support this program as the successor to the Collaborative.

The awardee planned to continue to advocate for its Community Care Center model by working on analyses that would support a future demonstration project. The University of Kansas Health System anticipated the need to secure waivers to depart from the standard CAH, rural health clinic, and FQHC models and test the feasibility, both financially and from a quality-of-care standpoint, of the transformational model. The awardee's staff noted that while three other states have proposed similar rural health models, they would seek to advance their model, as it was the furthest along in development.

D. Factors associated with the development of the payment model

The awardee identified few barriers to implementing its two new payment approaches. None of the barriers were significant. One barrier worth noting was gaining rural providers' attention, trust, and agreement to participate in the MSSP. Another was that some rural providers do not have the resources (e.g., staffing, technology) to participate in the MSSP.

With respect to the Community Care Center model that the awardee designed and hopes to demonstrate, significant barriers include (1) the inability to obtain waivers and (2) the analytic resources needed to support such a model. First, the awardee's representatives said obtaining waivers from CMS to pursue alternative payment models was one of the biggest barriers to implementation. Specifically, they need a waiver for CMS's "nearest appropriate facility" reimbursement rule for ambulance transports made pursuant to network-approved acute care protocols and a waiver for the three-day acute care stay requirement for skilled nursing facility benefits. Second, awardee leaders stated that were it not for the cooperative agreement they would not have been able to fund the analytics necessary to participate in new payment models. They asserted that rural providers in general are going to have that challenge.

> "These organizations do not have deep pockets, nor do they have tremendous local resources, nor are there analytics centers that they can turn to plug in their information and be able to begin to understand what that means. And so the lack of infrastructure to support that type of analytics for small rural providers has been very evident through the work that we've been doing."
>
> *—Program manager*

The awardee leaders said that their alternative payment model had been developed in collaboration with the Kansas Hospital Education and Research Foundation, a subsidiary of the Kansas Hospital Association, as well as the National Rural Health Association and the United Methodist Health Ministry Fund. They mentioned receiving advice and assistance from the Federal Office of Rural Health Policy, CMMI, and the implementation and monitoring contractor.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The University of Kansas Health System reported that all sites had signed agreements to continue the program after the end of the cooperative agreement, either through participation in the ACO or the PSO. Program services will continue to be funded with revenue from CCM and TCM services furnished to traditional Medicare beneficiaries, MSSP shared savings, and dedicated funding from the University of Kansas Health System. Awardee leaders felt optimistic about their ability to engage external partners to help sustain the program, such as payers and state officials. Although external partners had not finalized plans to help sustain the program, awardee leaders reported positive feedback during earlier meetings with payers and the current secretary of the Kansas Department of Health and Environment. Furthermore, awardee leaders expect to continue expanding the program to additional sites beyond the cooperative agreement. The awardee did not report plans to replicate its program.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, the University of Kansas Health System planned to sustain the program by using (1) revenue from CCM and TCM furnished to traditional Medicare

beneficiaries, (2) any shared savings received through the MSSP program, and (3) KUH funding for specific staff and services. The awardee was developing an application to meet the regulatory requirements to participate in the MSSP and had developed a budget to fund the program three years after the end of the cooperative agreement. KUH formed the new legal entity, the KCIC, to operate as an ACO participating in the MSSP. The awardee reported that the ACO would support the TCM and CCM program services as well as performance improvement activities such as the acute care protocols and performance reporting, feedback, and monitoring.

C. Implementing the SSR plan: progress and changes

Sustainability. KCIC will serve as the successor organization for the Consortium after the program funding ends and will continue to provide TCM and CCM services and performance improvement activities using the revenue generated from these services and shared savings to pay for health coaches, physician supervision, technology solutions, the eHealth Coaching contract, billing, compliance, and management. KCIC will support any sites that want to provide TCM or CCM services in establishing and maintaining their programs. By the end of the third program year, the awardee reported that all sites had signed agreements to continue participation in the program after the end of the cooperative agreement, either through participation in the ACO or the PSO. Awardee leaders believed the same program services would be provided under the sustained model, although possibly under a new, rebranded program name. Moreover, one awardee leader reported that most hospitals will continue using Avera, although a few may not due to financial constraints.

Program leaders at the University of Kansas Health System were hopeful about their abilities to engage external partners to help sustain the program, such as payers and state officials. The awardee had meetings with three managed care organizations (MCOs) scheduled in August 2017. Based on previous conversations with Blue Cross Blue Shield, the Kansas Medicaid program, and the MCOs, an awardee leader reported that payers generally had positive reactions to the program. “They were all excited and wanted to stay engaged,” the leader said. In the August 2017 meeting, the awardee planned to re-engage payers and start working out how to financially sustain the program. The awardee also met with the current secretary of the Kansas Department of Health and Environment in the third program year. The awardee leader and secretary discussed the department’s plans to create more Medicaid health homes and how KCIC could partner to become a Medicaid health home for Medicaid patients with chronic conditions while using the care models developed under this program.

Scalability. Awardee leaders expect to continue expanding the program to additional sites beyond the cooperative agreement. The awardee has developed program resources, including training and protocols, to support scaling. By the end of the third program year, sites in about half of the state’s counties had implemented the program. Although awardee staff are not actively marketing the program to new sites, growing participation in the MSSP and PSO and the positive reputation of the program have already generated interest from other rural providers.

Replicability. The University of Kansas Health System did not report plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

Program leaders reported being focused on sustainability from the start of the cooperative agreement. For example, the awardee adopted specific technology that was less costly instead of broad technology solutions that were more costly and would have been difficult to fund after the cooperative agreement ended. As the Collaborative's executive sponsor, leaders at the University of Kansas Health System approved of and supported the Collaborative's long-term sustainability plan. Moreover, awardee leaders believe that scaling the program through the MSSP will help support the program as new participants bring in revenue and shared savings. However, awardee leaders reported that the program's budget will be tight after HCIA R2 funding ends, particularly because TCM and CCM reimbursements do not cover overheard costs. Some sites experienced high overhead costs to meet the billing requirements for TCM and CCM services because they needed to implement and update enhanced EMR features. The awardee will continue to look for ways to operate the program more efficiently and for diverse sources of funding.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension that allows it to continue to enroll new beneficiaries. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in the University of Kansas Health System's program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicare beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: University of North Carolina

April 18, 2018

Julia Doherty (L&M Policy Research)
Abby Hoffman (L&M Policy Research)
Poonam Pardasaney (RTI International)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes. In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

The University of North Carolina at Chapel Hill (University of North Carolina) used HCIA R2 funding to implement the Better Back Care (BBC) program to increase the use of evidence-based care for acute, non-specific lower back pain (LBP) (Table I.1). The awardee designed

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The University of North Carolina received a one-year extension through August 31, 2018.

BBC to guide providers at each of 36 participating sites[3] in treating patients with acute LBP in the most conservative—yet appropriate—manner. The awardee aimed to improve care by emphasizing conservative treatments such as home exercises, physical therapy, and consultation with a pain psychologist or exercise physiologist, over invasive and costly, imaging, injections, and surgeries when appropriate. The University of North Carolina enhanced services provided for acute LBP patients by offering primary and specialty care providers a newly developed evidence-based checklist that functioned as a decision-support tool to support their care to participants with acute LBP. The awardee also enhanced care by hiring nurse care managers to enroll and provide participants with assistance including phone-based care management and monitoring, and support accessing appointments with program-affiliated specialists.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The University of North Carolina intended to encourage conservative, evidence-based treatment of acute LBP with its BBC program.
Major innovation	An LBP checklist, which was integrated into the electronic medical record (EMR) or provided on paper, that (1) prompted participating providers to follow an evidence-based treatment protocol for all patients presenting with new, acute LBP and (2) offered decision support.
Program components	• Care management • Evidence-based treatment for acute LBP • Health information technology • Shared decision making
Target population	Participants seen at BBC practices for the first time or within their first episode of care who met the following criteria: • Were seeking care for a new episode of LBP with a pain duration of less than 3 months and no clinical red flags • Had not seen a provider for back pain in the previous six months • Were 18 years or older • Were Medicare beneficiaries (fee-for-service (FFS) or Medicare Advantage) or Medicaid recipients • Spoke English or Spanish
Theory of change/ theory of action	The BBC program was intended to decrease costs and improve quality of care for patients with acute LBP through the following: • Encouraging participating providers to use an evidence-based checklist during the first office visit • Employing nurse care managers, who facilitate appropriate access to care, shared decision making, and coordination of care across primary and specialty care providers • Ensuring a referral network of specialists as well as other professionals—such as an exercise physiologist and pain psychologist—who provide conservative, evidence-based care

[3] The awardee recruited practices primarily from four medical groups including primary and specialty care providers from the University of North Carolina Faculty Physicians, the University of North Carolina Physicians Network, Piedmont Health Services (made up of community and rural health centers) and Advance Community Practice. These group practices also function as training sites for the University of North Carolina School of Medicine.

Table I.1 *(continued)*

Program characteristic	Description
Payment model	Rather than developing a payment model, the awardee encouraged adoption of the BBC care model throughout the University of North Carolina's health system and its affiliates where alternative payment models support the value orientation of BBC.
Award amount	$6,034,888
Effective launch date[a]	February 23, 2015
Program setting	Participating practices and in the community
Market area	Rural, urban, and suburban
Market location	Five-county region in central North Carolina
Target outcomes	• Increased patient satisfaction • Improved patient-reported general health and physical function • Decreased provision of care that does not conform to evidence-based guidelines

BBC = Better Back Care (BBC); LBP = lower back pain.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was partly effective in implementing its program by the end of the three-year cooperative agreement. We based this conclusion on six factors. First, the awardee achieved 101 percent of its adjusted enrollment target by the end of the cooperative agreement. Second, the awardee implemented the program at 36 provider practice sites, educated providers about how to use the LBP checklist and practice appropriate conservative care, and provided nurse care management support as intended. Third, the awardee recruited, hired, and trained qualified program staff in a timely manner and retained most staff throughout the cooperative agreement. Fourth, the awardee struggled to attain the involvement of providers at participating sites, many of whom did not use the evidence-based LBP checklist, conduct depression screenings, or distribute informational handouts in accordance with the BBC service delivery model. Fifth, the awardee experienced a less than full level of participant engagement; one-quarter of participants did not engage in initial calls with nurse care managers and 40 percent did not engage in follow-up calls with nurse case managers. Participants also frequently failed to attend appointments with the BBC exercise physiologist or pain psychologist to whom they were referred. Finally, while participating clinicians and other staff reported they believe the program had a positive effect on the delivery of care, when interviewed, program leaders said they wanted to wait for the results of their internal claims analyses and program evaluation before reaching conclusions about the program's success.

Impact evaluation. Due to too few treatment beneficiaries, we do not anticipate being able to conduct a rigorous impact analysis for the University of North Carolina. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. Rather than developing a payment model, the University of North Carolina encouraged adoption of the BBC care model throughout the University of North Carolina Health Care's clinically integrated networks and system (UNC system), where alternative payment

models support the value orientation of BBC. The awardee initiated negotiations with the management team of the Medicare Next Generation Accountable Care Organization (ACO) to use the BBC model throughout the ACO provider network, and potentially in value-based payment arrangements with commercial insurance companies. In addition, the awardee filed a recommendation with the International Classification of Diseases, Tenth Revision (ICD-10) coding and maintenance committee advising coding changes and enhancements to Medicare's Merit-Based Incentive Payment System to differentiate between acute and chronic LBP.

Sustainability plans. The University of North Carolina is working with the UNC system and its ACO to encourage sustainment of some BBC program activities. Specifically, the awardee reported that its leaders are working with the UNC system to add new decision-support features into its primary EMR system, designed to assist providers when treating patients with LBP and reinforce best practices. In the third program year, the UNC system also began exploring ways to scale the use of best practices when treating acute LBP throughout its provider network. The University of North Carolina piloted an audit and feedback method with one BBC practice, which will help the UNC system understand how to encourage best practices when treating acute LBP in system-affiliated hospitals and practices.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of clinicians on their perceptions of the effect of the program on the delivery of care. The clinician survey was fielded from March to June of 2017 to all 110 providers across the 36 implementing sites, and achieved a response rate of 67 percent overall, though not every respondent answered every question. Participating physicians were in practices providing primary, urgent, and emergency care, as well as specialty care provided in orthopedic and spine centers. We did not weight the survey to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partly successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

The University of North Carolina passively enrolled direct and indirect participants. Direct participants were individuals who were seen at the participating practices, sought care for a new episode of LBP with a pain duration of less than three months, had not seen a provider for LBP in the previous six months, were at least 18 years old, were Medicare or Medicaid beneficiaries, and spoke English or Spanish as a primary language.

The University of North Carolina changed its direct participant enrollment strategy during the first year of the cooperative agreement. At the outset, the awardee relied on participating providers to complete at least part of the LBP checklist with patients who presented with LBP and actively refer them to BBC for potential enrollment in the program. However, given the low number of providers who referred potential participants, awardee leaders set up a new process

wherein nurse care managers began to identify eligible participants without provider input. The nurse care managers first reviewed daily diagnosis code reports of LBP patients who saw participating providers to identify potentially eligible participants. Then, they reviewed the potentially eligible participants' medical records, including any completed or partially completed LBP checklists. If the nurse care managers found no disqualifying information during the medical record review, they enrolled the individuals into BBC as direct participants by entering them into the BBC customized care manager decision-support tool and patient database. BBC began this new enrollment process at the end of the first program year (October 2015) and continued to use this process throughout the cooperative agreement.

The University of North Carolina did not initially intend to serve indirect participants. However, in response to participating providers' desire to offer all patients at least some BBC services, regardless of their insurance status, the awardee began to enroll and count indirect participants in the third quarter of the program. Indirect participants met the same eligibility criteria as direct participants except that they were insured by a carrier other than Medicare or Medicaid. Indirect participants did not receive the full range of BBC services offered to direct participants. Instead, they attended office visits with a participating provider and were offered the opportunity to attend group exercise classes with the exercise physiologist; they did not receive support from BBC nurse care managers.

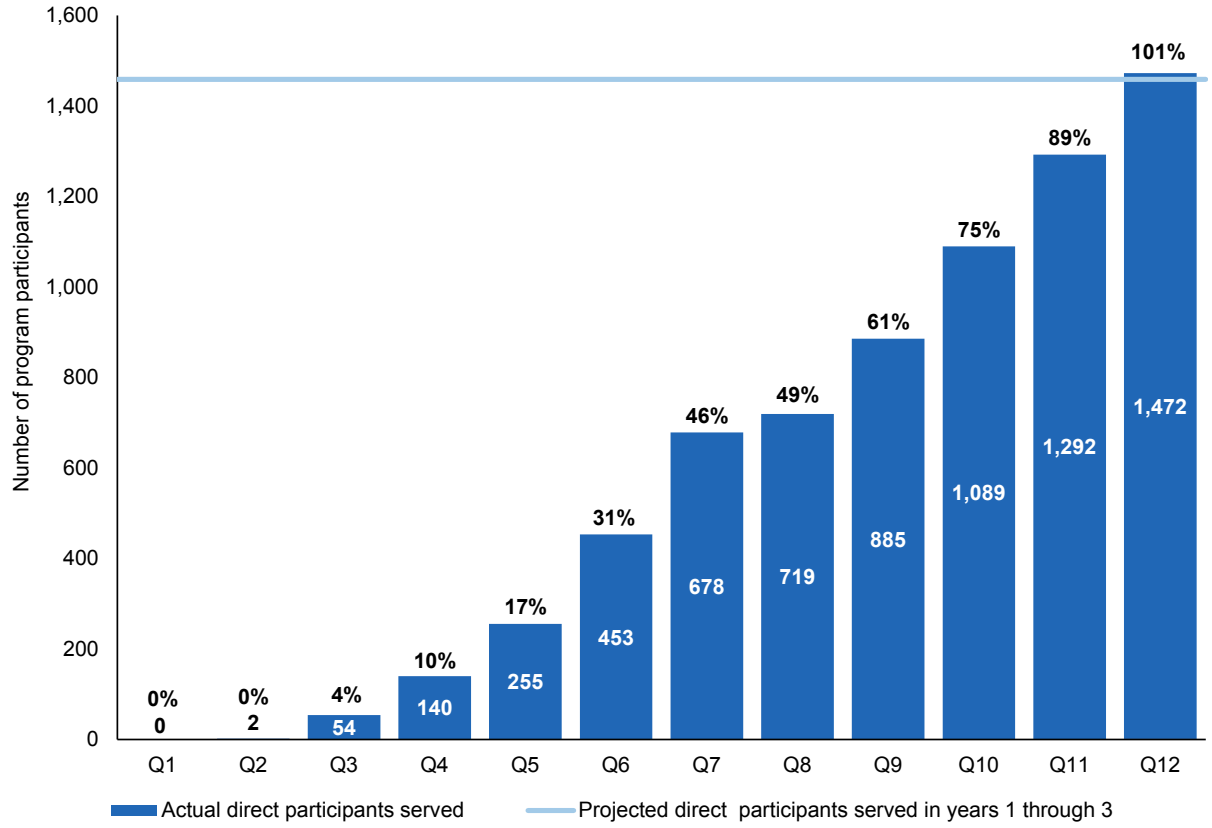
b. Evidence of enrollment effectiveness

Based on the awardee's revised enrollment projections made at the beginning of the third program year, the University of North Carolina achieved enrollment effectiveness. Overall, the awardee reported that it enrolled 1,472 direct participants from February 2015 (when it launched its program) through August 2017, representing 101 percent of its final three-year projection (Figure II.1). However, the University of North Carolina recalibrated its enrollment projections for direct participants downward two times during the three-year cooperative agreement. The awardee set its initial three-year enrollment target at 10,662 direct participants, but lowered the projection to 2,046 early in the second program year, and to 1,459 in the third program year. When measured against the three-year projection set at the end of Year 1 (enough time for the awardee to gain a better understanding of the number of eligible participants but before the awardee began planning for the end of the cooperative agreement), the awardee met 72 percent of its projection. The University of North Carolina recalibrated its direct enrollment goal in the second program year in response to learning in the first program year that it had significantly overestimated the pool of potential Medicare and Medicaid beneficiaries. During interviews in the first program year, awardee leaders reported that the original enrollment target was based on research conducted on commercially insured patients, rather than Medicare and Medicaid populations, which likely resulted in overestimating the prevalence of LBP in participating provider practices. During the third program year, the awardee reported further reducing their enrollment projections due to fewer patients than anticipated being enrolled through the ED.

The awardee also reported that it had served 280 indirect participants between February 2015 through August 2017, representing 69 percent of its three-year projection (Figure II.2). Like the direct participant targets, indirect participant targets also changed over time. The awardee noted the first three-year projection of 671 indirect participants in its fourth quarterly

report. The awardee increased its indirect target to 800 during the fifth program quarter, and then decreased its final indirect projection to 400 during the eighth program quarter.

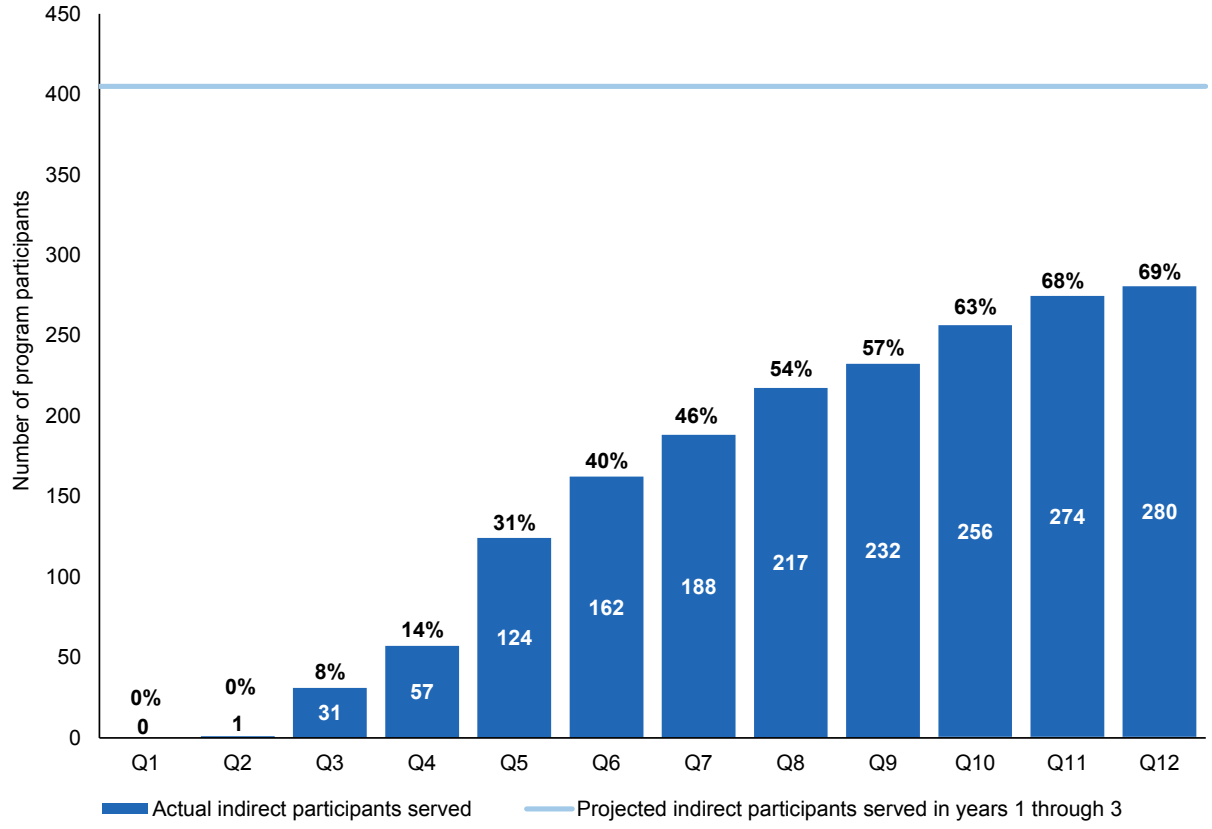
Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee lowered its direct participant enrollment target several times during the three-year cooperative agreement, from 10,662 at the beginning of Year 1 to 2,046 at the beginning of Year 2 to 1,459 during Year 3.

Figure II.2. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers. The awardee changed its indirect participant enrollment projection several times during the there-year cooperative agreement, from 671 indirect participants in Q4 to 800 in Q5 to 400 in Q8.

c. Barriers and facilitators associated with enrollment effectiveness

The University of North Carolina's progress in meeting its three-year enrollment goals was influenced by several factors. In addition to the enrollee's inaccurate enrollment projections at the onset of the cooperative agreement, the University of North Carolina identified two factors that hindered enrollment efforts: (1) a referral mechanism that was dependent upon technology availability and provider initiative and (2) nurse care manager turnover and short-term leaves of absence. In addition, we identified two factors that facilitated its enrollment effectiveness: (1) BBC leaders and staff's creativity and flexibility in problem solving and (2) the availability of data from the UNC system's EMR.

According to awardee leaders and staff, the University of North Carolina's initial approach to identifying potential participants depended on providers initiating the LBP checklist to generate a referral to nurse care managers. This significantly hindered enrollment efforts. This

referral mechanism required that (1) the checklist technology be available and easy to use on the program launch date at each participating site, and (2) that providers actively use the LBP checklist with every potentially eligible participant. Both requirements presented challenges. Awardee leaders said that because the UNC system's transition to a new EMR platform occurred simultaneously with BBC's implementation, the technical work of integrating the LBP checklist and referral mechanism into the new EMR platform fell behind more urgent system-wide EMR tasks. Integrating the LBP checklist and referral mechanism into the EMRs of practices outside of the UNC system was also complicated. Nurse care managers and providers reported that the EMR referral mechanism worked inconsistently, causing some providers to miss some potentially eligible participants. In addition to technological barriers, most interviewees across the three years of the cooperative agreement indicated that providers faced many competing demands on their time and attention and thus often did not prioritize using the LBP checklist or referring participants to the BBC program.

According to interviewees, nurse care manager turnover and short-term leaves of absence also hindered enrollment efforts. One care manager left BBC in the second program year, while another took several weeks of personal leave during the third program year. Due to the turnover and leaves of absence, staff fell behind in reviewing diagnosis code reports and medical records to identify potential participants. As a result, some potential participants were excluded from the program because the nurse care managers could not complete the initial call within two days after the patient's initial clinic visit, as required by the BBC service delivery model.

The awardee also experienced two facilitators of enrollment effectiveness. Awardee leaders and staff repeatedly demonstrated their ability to creatively problem solve and overcome barriers to enrollment. For example, to increase enrollment beyond participants referred by providers the awardee revised the care managers' role to include identifying potential participants by using daily lists of LBP diagnosis codes and reviewing medical record data. The University of North Carolina also expanded eligibility to include Spanish-speaking patients by hiring a part-time Spanish speaking nurse care manager to provide care management support. The awardee also increased the exercise physiologist's hours to full-time so she could help nurse care managers identify and enroll participants. The awardee was especially creative in addressing low enrollment by expanding the number and types of practices beyond primary care that it included in the intervention that cared for LBP patients. For example, the awardee accelerated and broadened its recruitment activity after leaders determined that patients with acute LBP tend to seek care from more diverse settings than were originally included in the BBC program, such as urgent care centers and EDs.

Finally, program staff repeatedly mentioned during interviews that access to data including ICD-10 codes that were available through providers' billing systems and EMRs facilitated participant enrollment efforts. Using these data, nurse care managers enrolled patients without direct referrals from providers and readily accessed EMR systems to confirm patient eligibility and communicate with providers about participants. BBC leaders also used data from EMRs to identify UNC system practices that saw a large number of acute LBP patients and to target those practices for recruitment.

2. Delivery of program services

a. Description of and changes to service delivery model

The BBC program was a new initiative by the University of North Carolina under the HCIA R2 cooperative agreement, and the awardee made no significant changes in the service delivery model over the course of the award. In designing BBC, the awardee hypothesized that encouraging providers to use a LBP checklist to guide the provision of evidence-based care would lead providers to recommend the most appropriate and conservative care options, and thus reduce use of more invasive and costly spinal imaging, injections, and surgeries among both direct and indirect participants. The awardee also expected that employing nurse care managers, who facilitated direct participants' access to care and participation in shared decision making, would guide participants to the most conservative, yet appropriate care. Together these intervention strategies, the awardee hypothesized, would result in (1) improvements in the quality of care for patients with LBP, (2) improvements in patients' self-reported health outcomes and satisfaction with care, and (3) decreases in the cost of care for patients with LBP. BBC's ability to complete many of the steps listed below was dependent on participants' willingness and ability to comply with provider and care manager treatment recommendations. The key steps in the BBC service delivery model are as follows:

1. BBC leaders supervised the embedding of the LBP checklist into EMRs or distributed paper copies of the checklist to participating practice sites. Paper copies were phased out after the first year of the cooperative agreement.
2. BBC program staff (including clinicians) and nurse care managers trained providers at each participating practice site about how to use the checklist and associated patient handouts when performing an LBP assessment.
3. Participating providers completed the evidence-based checklist tool (which included depression screening) with all patients presenting with acute LBP and documented checklist results in each patient's EMR.
4. Nurse care managers and the exercise physiologist entered descriptive data for direct participants from the EMR into the customized care manager decision-support tool and patient database developed by the awardee for all enrolled direct participants.
5. Nurse care managers contacted direct participants within two days of their provider visit, using the customized care manager decision-support tool and patient database to track participant contacts, guide treatment discussions with participants, and determine appropriate follow-up.
6. Nurse care managers contacted direct participants again two weeks after their initial office visit for LBP to follow up and provide additional support as needed. Follow-up could include reinforcing provider recommendations such as exercises to mitigate back pain. Care managers also worked with direct participants and providers to facilitate referrals to the BBC pain psychologist, exercise physiologist, back pain specialist, or to non-BBC providers such as physical therapists and specialists that BBC leaders identified as practicing appropriate conservative LBP care. BBC also offered indirect participants an opportunity to attend group exercise classes with the exercise physiologist; indirect participants did not receive additional support from BBC nurse care managers.

7. For interested direct participants, nurse care managers continued to check in via telephone, helping to coordinate care, and encouraging positive health behaviors such as exercise. This care coordination and support could continue for as long as nurse care managers deemed necessary to treat an episode of acute LBP, up to three months.

b. Evidence of service delivery effectiveness

Overall, the University of North Carolina partially achieved service delivery effectiveness. The awardee's achievements included: (1) providing nurse care management services in accordance with the service delivery model and (2) hiring, retaining, and training the staff it needed to deliver program services. However, the awardee struggled to attain the involvement of providers at participating sites, many of whom did not use the evidence-based LBP checklist, conduct depression screening, or distribute informational handouts in accordance with the BBC service delivery model. In addition, the awardee experienced less than full participant engagement; one quarter of participants did not engage in initial or follow-up calls with nurse care managers and more participants failed to attend appointments with providers to whom they were referred. Details are provided in the sections below.

Delivery of intervention services. The University of North Carolina was partly successful in delivering the BBC intervention services. According to awardee leaders and staff, the University of North Carolina educated providers about how to use the LBP checklist and practice appropriately conservative care, and provided BBC nurse care management support and access to the BBC exercise physiologist and pain psychologist to direct participants as intended. The awardee also offered group exercise classes to direct and indirect participants as intended.

The awardee successfully deployed nurse care managers and developed a customized care manager decision-support tool to support the delivery of appropriate and coordinated care to direct participants. The nurse care managers reported no problems in using the customized care manager decision-support tool and patient database, which prompted them to ask specific questions and provided assistance in suggesting evidence-based next steps. The nurse care managers described offering follow-up care to participants and communicating with providers to facilitate referrals as appropriate.

The BBC exercise physiologist and pain psychologist also provided follow-up care based on referrals from the nurse care managers and participating providers, as specified in the program model. As of August 31, 2017, the exercise physiologist had received referrals for 244 direct and indirect participants, and the pain psychologist had received 43 referrals for direct participants. To meet the needs of referred participants, the exercise physiologist held 48 small group classes in seven locations; group classes were open to direct and indirect participants.

The awardee reported, however, that a minority of direct and indirect participants attended initial provider office visits—which included the LBP checklist, depression screening, or use of informational LBP handouts—and that the percentage of participants that received these services decreased over time. That decrease is likely due to changing enrollment processes. Throughout most of the first program year, only patients whose providers identified them as having acute LBP and used the LBP checklist were eligible to become participants. In later years, the awardee enrolled participants based on diagnosis code list and medical record data review and conversations with patients to confirm they met the eligibility criteria, rather than based on

provider use of the LBP checklist. Corresponding with changes in enrollment processes, providers conducted core service delivery activities more frequently in the first program year and less frequently in later years. For example, the awardee reported to the implementation and monitoring contractor that providers' monthly LBP checklist completion rate for BBC participants ranged from 50 percent to 80 percent in the first program year, and was consistently less than 20 percent by 2017. Depression screenings and use of evidence-based handouts followed similar trajectories.

Staffing and training. The University of North Carolina successfully hired and retained staff at anticipated levels and by the twelfth quarter had nine full-time equivalent staff members. Key staff included program administrators, nurse care managers, and principal investigators. With the notable exception of a few months during the second program year when a care manager left the program, BBC leaders indicated that the staffing for BBC was appropriate and adequate to deliver BBC services as intended (though it was not always adequate for enrollment, as described previously). Program staff interviewed in the second and third years of the cooperative agreement reported that staff generally had an appropriate workload. Staff training was not fully developed when the first few BBC staff were hired; however, care management training was successfully developed and deployed with subsequent hires.

Recruitment and engagement of providers. The awardee successfully recruited 36 practice sites to participate in BBC, but struggled to engage the majority of individual providers in those practices. In the clinician survey, for example, 20 percent of respondents indicated that they were not at all familiar with the acute LBP checklist and 45 percent reported that they had never used it. Despite relatively low LBP checklist use, providers who completed the clinician survey reported generally being aware of the program, with most saying they were familiar with BBC nurse care management services (84 percent), access to the exercise physiologist (65 percent), and access to expedited specialty appointments (59 percent). In addition, multiple interviewees reported that some physicians, though a minority, were fully engaged and referred every patient with acute LBP to the program.

Awardee leaders also reported that the specialty providers to whom they referred BBC participants were generally engaged in the program and often expedited appointments for participants, as required by BBC.

Engagement of program participants. The awardee was partially successful in engaging participants. According to the University of North Carolina's twelfth quarterly report, three-quarters of enrolled direct participants engaged in the initial call with the nurse care manger and 60 percent participated in the two-week follow-up call. In interviews, however, multiple staff indicated participants frequently missed appointments with the exercise physiologist or other specialists arranged for them by BBC care managers.

c. Barriers and facilitators associated with service delivery effectiveness

The University of North Carolina's ability to deliver intervention services was influenced by several factors. Awardee leaders, program staff, and providers highlighted several primary barriers and facilitators to service delivery effectiveness. Looking over the full three-year cooperative agreement, five main facilitators emerged: (1) ongoing education to, and communication with, providers in participating practices, (2) use of provider profile performance

measure reports, (3) building a supportive network around providers to increase their comfort with conservative LBP care, (4) decision-support tools, and (5) participants reporting their positive experiences to their providers. Two barriers mentioned across the life of the project included: (1) competing demands on providers' attention and time and (2) difficulty integrating the LBP checklist into the EMRs.

First, BBC leaders and staff said that providing ongoing education to, and maintaining communication with, participating sites facilitated provider participation in the program. Education and communication took on multiple forms, including: initial onsite onboarding and training; ongoing onsite education from nurse care managers and other leaders (particularly in the first two years); and establishing champions within practices to encourage continued provider focus on BBC. These efforts helped to ensure that most providers, even those who rotated through clinics for short periods of time, were aware of and participated in the intervention.

A second factor that program leaders felt fostered service delivery was the use of provider profile reports that documented practice and provider use of selected evidence-based LBP practices, such as ordering spinal images or injections, prescribing narcotic pain medications, and using the LBP checklist. Program leaders said that the provider profiles allowed providers to see their own utilization rates and make comparisons against other practices. This, in turn, engendered competition and discussion among some of the providers about how to improve use of evidence-based practices and increased pressure on providers to reduce inappropriate testing.

Third, program staff reported that building a supportive network around providers as an integral part of the BBC program fostered use of evidence-based LBP practices. Specifically, program staff said that providers felt more comfortable and had more confidence in their ability to treat acute LBP conservatively because they trusted the clinical knowledge and expertise of the nurse care managers and knew the managers were there to provide additional support to participants if needed. In the final site visit, the BBC clinical leader said that the decision to hire nurses as care managers rather than social workers was the right one. The leader appreciated that nurse care managers could (1) deal with clinically complex participants and identify potentially troublesome symptoms, (2) communicate effectively with the physicians about any concerns and potential referrals, and (3) refer participants to additional providers if needed. In addition to the nurse care managers' expertise, multiple interviewees also noted that the program's access to a specialty referral network and the ability of nurse care managers to facilitate expedited appointments for participants strengthened providers' willingness to use conservative evidence-based practices.

Decision-support tools, including the LBP checklist and the customized care manager decision-support tool and patient database, also facilitated service delivery by focusing providers and nurse care managers on the delivery of evidence-based care. One physician leader indicated that the decision support embedded within the LBP checklist improved providers', especially medical residents', understanding of evidence-based conservative care, resulting in fewer unnecessary orders for imaging and injections. Multiple interviewees also reported that the customized care manager decision-support tool and patient database was a major facilitator throughout the project because it created a database of measurable patient characteristics, prompted nurse care managers to make follow-up calls, and provided them with the evidence-based decision support needed as they helped participants make optimal choices about their care.

Finally, two BBC leaders reported that when providers heard from participants about their positive experiences with BBC, the providers seemed more likely to participate actively in the program. Hearing positive participant experiences, they said, helped providers recognize the value of BBC services, and inspired them to use the LBP checklist more frequently and refer additional patients to BBC.

Perhaps the most significant barrier to the effectiveness of service delivery (which interviewees continuously reported during site visits each program year) was insufficient provider engagement due to competing demands on their attention and time, and difficulty changing long-established practice patterns. Interviewees said that providers were "inundated" with quality metrics, EMR flags, and other quality improvement processes focused on more prevalent conditions, and therefore often lacked the time or capacity to focus on BBC. Over 40 percent of respondents to the clinician survey indicated that one reason they might not complete the LBP checklist would be because they did not have time to walk through it with patients. Adding ED providers to the intervention introduced a new set of competing demands. For example, in its tenth quarterly report, the awardee said that ED providers were busy and experienced high patient turnover, leading them to forget to complete the LBP checklist. During the first program year, interviewees also said that some providers saw as few as two acute LBP patients per month, which made building the habit of completing the checklist more challenging. Resistance to changing well-established processes also made engaging providers difficult. One leader said, for example, that some specialists, such as orthopedists, in trying to be efficient with their time, sometimes ordered imaging before they assessed the patient—a practice that went against the tenets of BBC.

A second ongoing barrier to service delivery was the awardee's inability to fully integrate the LBP checklist into all participating practices' EMR systems and the work required to accommodate EMR system changes. For example, the UNC system was implementing a new EMR across the health care system and associated providers at the same time BBC was initiated. Because there were multiple types of EMR system configurations (e.g., between primary care and the ED), difficulties implementing the LBP checklist persisted even after initial EMR implementation was finalized. In addition, EMRs at practices outside the UNC system also faced early glitches in LBP checklist functionality. The awardee continually worked to overcome EMR integration barriers throughout the three-year cooperative agreement. A BBC leader said that while the LBP checklist was never fully integrated into every EMR system, by the third program year BBC program staff were able to implement a workaround that streamlined use of the EMR systems in practices where the most user friendly version of the LBP checklist was not embedded.

C. Assessment of perceived program effects on the delivery of care and outcomes

Overall, the BBC staff and providers we interviewed believed that the program had a positive impact on the delivery of evidence-based LBP care. The perceptions of clinicians who completed the clinician survey, however, varied in terms of whether and how the program influenced the delivery of care. For example, only 63 percent believed that BBC improved the quality of care for participants and 43 percent said that the program positively impacted

providers' ability to respond in a timely way to patient needs. Most providers surveyed, however, believed that the program had a positive impact on care coordination (Table II.2).

Table II.2. Clinician perceptions of BBC program effects on care

The BBC program had a positive impact on:	Percentage of clinicians who agreed with the statement (n=74)
The quality of care and services you provide to participants	63
Your ability to respond to participant needs in a timely way	43
Your ability to provide care or services that are responsive to participant preferences, needs, and values	42
Access to care or services for all participants	63
Achievement of participants' health goals	61
Participant satisfaction	64
Participant quality of life	59
Care coordination	82

Source: HCIA R2 evaluation survey of awardee's clinician survey, March to June of 2017.

Note: This survey had 74 respondents overall. A small number of responses (three to five) were missing for some survey items.

Most interviewees also said that the program positively influenced BBC's desired outcomes of improved patient-reported health and satisfaction with care and decreased rates of unnecessary spinal images and injections. However, some program leaders said they wanted to wait for the results of their internal claims analyses and program evaluation before reaching conclusions about the program's success. In terms of patient satisfaction, the awardee's internal analyses demonstrated achieving its goal of having 90 percent of participants respond positively to patient satisfaction and shared decision-making questions during follow-up calls conducted by nurse care managers by the last quarter of 2016. In addition, one leader said that patients reported positive experiences with BBC to their primary care providers (PCPs). Finally, more than half of providers who completed the clinician survey reported that the program positively impacted participant satisfaction (64 percent) and participant quality of life (59 percent) (Table II.2).

The awardee's internal, preliminary analyses also showed reductions in non-evidence-based service utilization among participants. For example, the awardee reported that imaging rates at the first clinical visit for LBP were maintained at a fairly steady 21 percent between the tenth and twelfth quarters. This was down from a high of more than 30 percent in early 2016. The awardee reported in interviews and supplemental documents, however, that imaging rates generally remained higher across the course of the cooperative agreement for ED and spinal specialists than for PCPs. Nearly half of providers who completed the clinician survey (43 percent) reported that they observed lower imaging rates as a result of their practice's participation in BBC and use of the LBP checklist. BBC leaders we interviewed also said that the program likely decreased use of unnecessary imaging and injections, but expressed concern that these improvements would not be retained after BBC staff stopped monitoring and reporting on LBP checklist and imaging use, and no longer offered care management and support services tailored to LBP patients.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

Several factors may limit the ability of an impact analysis to fully detect the outcomes of BBC model intervention, to include potential participants either not being enrolled or not receiving the full intervention as originally intended once identified as participants. First, an unknown number of potentially eligible participants were not enrolled in the program for a number of reasons; this could mean that the participants that were enrolled were not generally representative of the target population. The reasons some potentially eligible participants were not enrolled include: (1) providers faced competing priorities and many did not prioritize use of the LBP checklist or referring patients to the program, resulting in some potential participants being missed during the first year of the cooperative agreement when the awardee was relying on providers to identify and refer potential participants, and (2) once care managers became responsible for identifying potential participants through the use of daily diagnosis reports, care managers sometimes fell behind in reviewing reports, EMR data and contacting potential participants within the required two days following the initial visit due to staff absences or turnover.

Second, not all participants received the full gamut of intervention services as originally intended, which lessened the potential influence of the intervention. For example, given that many providers did not use the checklist (45 percent of survey respondents reported never using it), there was no guarantee that participants actually received the core intervention services—use of the LBP checklist, depression screening, and informational handouts. In addition, when nurse care managers called participants to offer additional support after the initial provider visit, care managers were unable to reach a quarter of the participants. While less than full participant engagement with nurse care managers is unlikely to affect utilization driven by provider behavior during initial office visits, such as imaging and injection rates, it may have negatively affected the awardee's ability to improve patient-reported outcomes such as reductions in pain or satisfaction with care. These factors in combination likely mitigated the potential impact of the program.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of the University of North Carolina's BBC program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of December 1, 2017: University of North Carolina

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	616[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	7,366
Likelihood of all-cause hospitalizations	4,503
MDE sample size requirement to detect 20% effect	
Total expenditures	1,842
Likelihood of all-cause hospitalizations	1,126
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	We do not expect to select a comparison group due to small sample size and a high variance in outcomes
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Too few treatment beneficiaries for primary outcomes
Survey data for treatment group that will be analyzed	Clinician survey
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We do not expect to conduct a rigorous estimate of the impact of the University of North Carolina's BBC program. The projected sample is too small relative to the variance in outcomes to detect an effect of 20 percent on Medicare spending. Furthermore, as noted above, some beneficiaries who appear in the finder file may have previously received treatment for back pain. These beneficiaries were probably enrolled in the program because there was no indication of prior treatment in the university's EMR. However, our analysis of Medicare claims did identify this earlier treatment.

B. Characteristics of Medicare and Medicaid participants at baseline

The University of North Carolina began to enroll Medicare and Medicaid beneficiaries in the BBC program in February 2015. For the purpose of our evaluation, the treatment group consists of adult beneficiaries (age 18 years and older) in Medicare fee-for-service (FFS), Medicare Advantage, or Medicaid who were enrolled in the BBC program on or before May 31, 2016, according to lists from the awardee. The enrollment date was defined by the awardee as the date of the initial provider visit for LBP. For this report, we used the finder file received on November 11, 2017, which contained 678 unique beneficiary records.[4] A total of 62 records were dropped because Medicare was not the primary payer, because the beneficiary was not enrolled in both Part A and B at the time of program enrollment, or because the beneficiary was enrolled in Medicare Advantage—leaving 616 unique beneficiary records available for analysis.

For the purpose of presenting baseline characteristics in this report, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS (both Parts A and B) with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who had met all evaluation criteria for a period of at least 90 days within the baseline year (the 365 days immediately before their enrollment). The calendar period covered by the baseline quarters is based on the enrollment date for each participant and therefore varies by participant. In addition, the beneficiaries must have enrolled in the awardee's program on or before May 31, 2016, the last time we updated baseline characteristics until we receive the final finder file, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. These requirements, along with the exclusion of beneficiaries who did not have an identification number and those who could not be matched, eliminated 325 beneficiaries from the sample.

For this report, 228 Medicare beneficiaries met the above eligibility criteria and were included in the analysis of demographic and health status characteristics, as shown in Table III.2. Just over half of them (53 percent) are age 65 to 74, whereas 24 percent are 75 to 84, and 18 percent are younger than 65. Most participants are female (57 percent) and white (75 percent); 23

[4] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

percent are black. Twenty-eight percent of participants were originally enrolled in Medicare because of a disability. Fewer than 0.5 percent were originally enrolled because of end-stage renal disease (ESRD). Twenty percent of participants are dually eligible for Medicare and Medicaid, although in some cases Medicaid benefits may be restricted to the payment of co-insurance and deductibles. The average hierarchical condition categories (HCC) risk score of participants (0.95) is less than the average for Medicare FFS beneficiaries nationwide (1.0). Seventy-five percent of participants have an HCC score of 1.12 or lower, and 50 percent have a score of 0.71 or lower. Thus, BBC program participants are slightly healthier than the average Medicare FFS beneficiary.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the University of North Carolina's program through May 31, 2016

	All participants (N = 228)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	42	18
65 to 74	121	53
75 to 84	55	24
85 and older	10	4
Gender		
Female	129	57
Male	99	43
Race		
White	170	75
Black	52	23
American Indian, Alaska Native, Asian/Pacific Island American, or other	2	0.88
Hispanic	3	1
Original reason for Medicare eligibility		
Old age and survivor's insurance	163	71
Disability insurance benefits	64	28
ESRD[a]	1	0.44
Hospice[b]		
Medicare/Medicaid dual status, percent dual[b]	45	20
HCC score[c]		Statistic
Mean		0.95
25th percentile		0.5
Median		0.71
75th percentile		1.12

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Table III.2 *(continued)*

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the beneficiary made an initial visit to a physician for acute, nonspecific low back pain. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category.

Consistent with the mean HCC risk score, which was slightly below the national average, participants did not have high Medicare expenditures or high rates of service use in the 365 days before enrollment. Table III.3 shows baseline expenditure and health care utilization data for a common set of measures, including the four core measures from the Center for Medicare & Medicaid Innovation. Given that the University of North Carolina intends to reduce overall Medicare spending primarily by reducing the use of imaging and injections, findings from our analysis of baseline cost and utilization data for LBP-related imaging, injections, and provider visits are included in this report.

The following six measures increased in the fourth baseline quarter: (1) number of outpatient ED visits, (2) percentage with an outpatient ED visit, (3) percentage with a 30-day readmission among all discharges, (4) percentage of participants with a readmission among all participants, (5) percentage of participants with LBP-related imaging, and (6) percentage of participants with an LBP-related provider visit. Although the awardee excludes beneficiaries with a history of LBP in the past six months, the presence of LBP-related imaging and provider visits in the last two quarters may represent imaging and provider visits in the outpatient ED setting, which would not disqualify a beneficiary from enrolling in the program. In the last baseline quarter, 40.3 percent of participants who had an LBP-related provider visit also had an outpatient ED visit.

We did not note a pattern from one baseline quarter to the next in the remaining expenditure and utilization measures. We will continue to monitor baseline expenditures and utilization each quarter to determine whether different patterns emerge as the sample size increases.

We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments in total and by major types of services. The total average PBPM Medicare payment of $576 during the baseline year was much lower than the national average of approximately $792.[5] Average PBPM Medicare payments for acute inpatient ($220), physician ($164), and outpatient ($150) services were the largest drivers of the total cost of care.

During the baseline year, the rate of acute care hospitalizations was 201 per 1,000 Medicare FFS beneficiaries per year, whereas the rate of ED visits not resulting in a hospitalization was

[5] See the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

407 per 1,000 beneficiaries per year. Fourteen percent of participants had at least one hospitalization during the baseline year, and 24 percent had at least one ED visit. The rate of ED visits per 1,000 beneficiaries per year was higher in baseline quarter 4 (597) than in baseline quarters 1 through 3 (272 to 404). Twelve percent of participants had at least one ED visit during baseline quarter 4. The rate of primary care visits in any setting in the baseline year was 5,292 per 1,000 Medicare FFS beneficiaries per year, and the rate of specialist visits in any setting was 8,087 per 1,000 beneficiaries per year.

In the baseline year, 8 percent of all hospital discharges in the treatment group were followed by a readmission in the 30-day post-discharge period. At the participant level, however, only one percent of all beneficiaries had a hospitalization with a readmission in the 30-day post-discharge period. The percentage of participant-level hospital discharges with a 30-day readmission in the baseline year was lower than the national average for Medicare beneficiaries (18 percent).[6] Readmissions were noted only in the third and fourth baseline quarters.

We examined three awardee-specific measures in the baseline year: the rate of participants with LBP-related imaging, injections, and provider visits. Thirty-five percent of all participants had at least one LBP-related provider visit, whereas 13 percent of all participants had LBP-related imaging, and one percent received LBP-related injections. The percentage of participants with LBP-related imaging and provider visits was highest in baseline quarter 4, compared with quarters 1 through 3. Although the presence of LBP-related imaging and provider visits in the last two baseline quarters appears to run counter to the inclusion criterion of no LBP-related history or provider visit in the past six months, as previously stated, 40.3 percent of participants with an LBP-related provider visit in the last quarter also had an outpatient ED visit, which is permitted by the awardee. In addition, LBP-related imaging may reflect imaging that took place during outpatient ED visits.

[6] See the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the University of North Carolina's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	228	214	220	228	228
Average Medicare expenditures PBPM[a]					
Total	576 (121)	464 (81)	669 (156)	571 (119)	484 (98)
Acute inpatient	220 (83)	110 (45)	291 (124)	241 (97)	160 (75)
Inpatient other[b]	1 (1)	2 (2)	3 (3)	0 (0)	0 (0)
Outpatient[c]	150 (23)	146 (30)	177 (54)	145 (32)	123 (19)
Physician services	164 (14)	155 (16)	169 (18)	157 (14)	165 (18)
Home health	16 (6)	17 (9)	14 (8)	16 (7)	16 (9)
Skilled nursing facility	10 (17)	16 (15)	0 (0)	0 (0)	5 (5)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	15 (5)	18 (11)	15 (6)	11 (5)	15 (4)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	201 (95)	136 (56)	202 (64)	215 (110)	158 (57)
Outpatient ED visits	407 (88)	272 (102)	404 (146)	287 (120)	597 (125)
Observation stays	69 (17)	58 (33)	55 (31)	125 (46)	35 (25)
Primary care visits in any setting	5,292 (296)	5,002 (433)	5,670 (465)	5,159 (471)	5,284 (389)
Primary care visits in ambulatory settings	4,848 (273)	4,476 (356)	5,138 (417)	4,729 (385)	4,986 (365)
Specialist visits in any setting	8,087 (616)	7,746 (885)	8,367 (856)	8,026 (739)	7,953 (685)
Specialist visits in ambulatory settings	7,131 (562)	7,045 (794)	7,211 (780)	7,094 (700)	7,022 (617)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	14 (2)	3 (1)	5 (1)	5 (1)	4 (1)
Percentage with an outpatient ED visit[e]	24 (3)	5 (2)	7 (2)	7 (2)	12 (2)
Percentage with an observation stay[f]	7 (2)	1 (1)	1 (1)	3 (1)	1 (1)
Percentage with a 30-day readmission among all discharges	8 (4)	0 (0)	0 (0)	13 (13)	14 (10)
Percentage of participants with a readmission among all participants	1 (1)	0 (0)	0 (0)	0.4 (0.4)	1 (1)
Awardee-specific measures					
Percentage of participants with LBP-related imaging	13 (2)	2 (1)	2 (1)	4 (1)	7 (2)
Percentage of participants with LBP-related injection	1 (1)	0.5 (0.5)	0.5 (0.5)	1 (1)	0.4 (0.4)
Percentage of participants with LBP-related provider visit	35 (3)	10 (2)	10 (2)	13 (2)	21 (3)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Rather than developing a payment model, the University of North Carolina encouraged adoption of the BBC care model throughout the University of North Carolina Health Care's clinically integrated networks and system (UNC system), where alternative payment models support the value orientation of BBC. The awardee initiated negotiations with the management team of the Medicare Next Generation Accountable Care Organization (ACO) to use the BBC model throughout the ACO provider network, and potentially in value-based payment arrangements with commercial insurance companies. In addition, the awardee filed a recommendation with the ICD-10 coding and maintenance committee advising coding changes and enhancements to Medicare's Merit-Based Incentive Payment System (MIPS) to differentiate between acute and chronic LBP.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: (1) the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor and (2) narratives and payment model reports the awardee submitted to the implementation and monitoring contractor. We did not interview the awardee about its payment model because it did not advance from the development stage to active negotiations with one or more payers.

B. Description of the payment model

The University of North Carolina originally proposed developing an episode-based bundled payment. That approach was abandoned in the third year of the cooperative agreement due to multiple considerations within the UNC system and the challenges related to data described below. Rather than developing a payment model, the University of North Carolina encouraged adoption of the BBC care model throughout the UNC system's clinically integrated networks, where alternative payment models support the value orientation of BBC. The awardee initiated negotiations with the management team of the UNC system's Medicare Next Generation Accountable Care Organization (ACO) to use the BBC model throughout the ACO provider network, and potentially in value-based payment arrangements with commercial insurance companies. The awardee's goal was for the UNC system and ACO clinicians who bill for any service related to acute LBP to identify and document when their patient presented with a new episode of LBP and be assessed for adherence to the BBC evidence-based model using quality measures approved by the ACO and its medical groups. The awardee also believed that Medicare's chronic care management reimbursement could support efforts to ensure patients with acute LBP continue to receive care management services, although the care managers would not be dedicated to exclusively serving LBP patients.

The University of North Carolina also proposed ICD-10 coding changes and enhancements to Medicare's MIPS to support high quality care for acute LBP patients. The awardee maintained that coding changes could support efforts to integrate LBP-related metrics and reporting into system-wide payment models. The recommended changes to MIPS could allow providers to track and receive credit within the MIPS for delivering care consistent with the BBC model. The proposed coding and changes in quality metrics include:

- Changes to current Medicare codes to allow a provider to identify and document when a patient presents with a new episode of acute LBP. Recommended changes would add codes for duration including unspecified, acute (< 4 weeks), subacute (> 4 weeks but < 12 weeks), and chronic (> 12 weeks) for LBP (M54.5), dorsalgia, unspecified (M54.9), segmental and somatic dysfunction of lumbar region (M99.03), other intervertebral disc degeneration, lumbar region (M51.36), radiculopathy, lumbar region (M54.16), radiculopathy, lumbosacral region (M54.17), and intervertebral disc disorders with radiculopathy, lumbosacral region (M51.17).
- Modifications and additions to existing MIPS quality measures to increase focus on encouraging providers to deliver high quality care for acute LBP. Specific recommendations include modifying the existing measure "Use of Imaging Studies for Low Back Pain" to include patients over age 50 but exclude patients with red flags, and adding measures for opioids, use of spine-related injections, use of low back surgery, advice on activity, screen for red flags, and screen for risk of chronic back pain. Finally, the awardee recommended creating a new MIPS specialty measure set for neurosurgery and key specialists to include the LBP-related measures described above.

The awardee suggested that once providers could code distinctions between acute and chronic LBP in EMRs and are incented to consider the refined quality metrics, use of interventions developed as part of the BBC, such as nurse care management, improved care coordination, access to physical/exercise therapy, and an EMR checklist would be more easily encouraged and more frequently implemented.

C. Status of developing the payment model

The awardee initiated conversations with medical management of the UNC system ACO to encourage the delivery of appropriate conservative care for acute LBP using the BBC model throughout the ACO provider network (rather than identifying a payment specific to the BBC care model). The BBC approach to care delivery for acute LBP will also be incorporated into discussions about future agreements as part of the University of North Carolina's overall care management approach and practice guidelines as a system. In 2017, BBC filed a recommendation with the ICD-10 coding and maintenance committee advising enhancements to ICD-10 coding to differentiate between acute and chronic LBP.

D. Factors associated with development of payment model

> "Patients with new onset [LBP] seek initial care from many different types of providers and provider settings … This pattern leads to a diffusion of care settings and smaller than expected panel sizes. These small panel sizes limit the potential choice of alternative payment model designs."
>
> —Awardee leaders

The University of North Carolina faced several challenges in developing the originally proposed episodic bundled payment model and care management fee specific to the acute LBP population. The inability to obtain Medicaid claims data prior to the end of the cooperative agreement was a barrier for evaluation and payment model development, as approximately 30 percent of participants were Medicaid patients. The state Medicaid agency was in negotiations to reestablish a data-use agreement with the UNC system to share de-identified patient claims data, but the process stalled within the agency. In addition, the awardee

was beginning to calculate an overall cost savings for the remaining participants experiencing the intervention near the end of the cooperative agreement period, but it took a long time to accumulate sufficient data to allow for claims analyses of the results given the slow enrollment and relatively small number of patients presenting with acute LBP. Finally, according to the awardee, analyses of cost patterns showed small sample sizes due to diffusion of acute LBP patients across multiple providers and provider types. This diffusion yielded sample sizes that were inadequate to effectively measure outcomes and risk necessary to develop alternative payment mechanisms. The awardee indicated they have had no success in getting other North Carolina payers to engage in discussions about the BBC payment model given the small size of the population for payers to carve out separately from the University of North Carolina's broader payment discussions.

These findings and potential alignment of the BBC program with MIPS led the awardee to propose coding change recommendations and pursue other alternative payment strategies. The awardee maintained that benefits to integrating BBC into other models outweighed those associated with creating a separate payment model. In addition, the awardee maintained that this approach offered administrative simplification, increased likelihood of adoption by providers of evidence-based medicine and associated quality metrics at scale, and recognizes the value of addressing acute LBP treatment as an important aspect of a patient's care needs.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The University of North Carolina is working with the UNC system and its ACO to encourage sustainment of some BBC program activities. Specifically, the awardee reported that its leaders are working with the system to add new decision-support features into its primary EMR, designed to further assist providers when treating patients with LBP and reinforce best practices. In the third program year, the UNC system also began exploring ways to scale the use of best practices when treating acute LBP to more affiliated hospitals and practices. The University of North Carolina piloted an audit and feedback method with one BBC practice, which will help the UNC system understand how to encourage best practices when treating acute LBP in system-affiliated hospitals and practices.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, the University of North Carolina had started planning how to engage internal and external stakeholders to sustain the program. Internally, the awardee embedded the LBP checklist into the EMRs of implementing practices so that providers could access the checklist in the EMRs after the cooperative agreement ended. Program leaders

conducted outreach and education efforts to increase the use of program tools among program staff, and made plans to approach UNC system leaders about implementing and sustaining a modified version of the program across the entire system. Externally, the awardee had multiple discussions with one larger payer about developing an alternative payment model.

C. Implementing the SSR plan: progress and changes

Sustainability. As of the third program year, the University of North Carolina expected to sustain some parts of the program. The awardee reported that although the LBP checklist will remain in the EMR systems for BBC practices and practices employed by the UNC system, providers will not be required to use it. To facilitate decision making, the awardee instead planned to implement pop-up reminders, or EMR flags, in the UNC system's primary EMR system. The awardee reported that it has been developing algorithms for the EMR flags with input from PCPs and specialists across the system. Examples of EMR flags include if the provider should order imaging soon, the typical management for a type of patient, what to look for when referring patients to physical therapy, or the communication needs between PCPs and specialists. The awardee also prepared a resource describing best practices for acute LBP intended to be shared with practices within the UNC system that are most likely to encounter patients with LBP.

Awardee leaders also expected that some aspects of the BBC care management program would be sustained. A number of UNC system practices function as medical homes and provide care management services, so for those practices, existing care managers can incorporate BBC evidence-based guidelines for LBP education with patients as necessary. However, one awardee leader noted that the retroactive care management approach provided by PCPs is inherently different from the proactive nurse care management provided by BBC.

Scalability. In the third program year, the UNC system began piloting an audit and feedback method with one BBC practice to potentially expand the use of best practices when treating acute LBP to more affiliated hospitals and practices. The piloted audit and feedback method was similar to that used in the BBC program, and was designed to test how the UNC system might benefit from sharing information with providers about their ordering patterns and the importance of practicing evidence-based care for acute LBP. The UNC system has the ability to reach 10 hospitals and 3,400 target practices in its system to encourage best practices.

Replicability. The University of North Carolina did not report plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

The awardee reported that proposed EMR flags and integrating the LBP checklist in the system and ACO participation are designed to facilitate sustainment of some program activities. Awardee leaders believed that the EMR flags will help sustain an important part of the program by encouraging providers using that EMR system to consider best practices when making decisions about patient care. One respondent noted that the awardee's previous experience creating and successfully sustaining other types of EMR flags in the UNC system makes the BBC EMR flags more likely to be sustained. Second, awardee leaders also noted that participation in the ACO would allow the EMR flags and LBP checklist to be rolled out to all practices. Finally, the awardee reported that while the flags will be implemented in the EMR

used by employed UNC system providers, it may not be possible to incorporate such EMR flags into other BBC practices' systems because they function on different EMR systems that the UNC system does not own.

As the end of the cooperative agreement approached, the awardee speculated about how the UNC system could continue to offer the services of an exercise physiologist, but encountered multiple challenges. First, the services provided by the exercise physiologist are not currently reimbursable by CMS. Moreover, one interviewee noted it would be difficult to incorporate a new position such as the exercise physiologist into the UNC system, unless the position is "spread across a very large number of practices to reduce the cost per practice."

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Regents of the University of California at San Diego

April 18, 2018

Stephanie Peterson (Mathematica Policy Research)
Theresa Feeley-Summerl (Mathematica Policy Research)
Jim Reschovsky (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background and purpose of the HCIA R2 initiative

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.

The federal evaluation of HCIA R2 has two goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

The Regents of the University of California at San Diego used funds from HCIA R2 to implement the Heart Attack and Stroke Free-Zone (HSF-Z) program, an effort to improve care for patients who are at an elevated risk for cardiovascular disease. Key program characteristics are shown in Table I.1. The goals of HSF-Z are to reduce the incidence of heart attacks and

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

strokes in San Diego County, thereby decreasing their associated health care costs and mortality rates. HSF-Z is led by the University of California at San Diego in partnership with 10 San Diego–area medical groups, which handle the majority of all medical care provided in San Diego.[3]

Each participating medical group identified and recruited eligible patients (see target population in Table I.1) into the project, with support and materials from staff at the University of California at San Diego. The target enrollment goal for all three years of the program was 3,800 participants.

The University of California at San Diego intended to improve participants' health by raising their awareness of cardiovascular risk factors; introducing evidence-based medications; and giving them supportive, ongoing health coaching. Health coaches worked with physicians to ensure that participants were put on appropriate, evidence-based medication bundles for hypertension, diabetes, and other conditions that may raise the risk of major cardiovascular events such as strokes and heart attacks. They also educated participants about positive health practices and engaged them in their own care to enhance compliance with the drug regimens and to encourage lifestyle changes. The HSF-Z program was supplemented by a community-wide effort to educate people about the risks for cardiovascular disease (the Be There program) and a long-standing effort to educate physicians (the University of Best Practices).

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Reduce the incidence of heart attacks and strokes in San Diego County, along with their associated health care costs and mortality rates
Major innovation	Provide supportive, ongoing health coaching to participants, and introduce and provide education on evidence-based medication bundles to patients and providers
Program components	Care management through health coaches, including medication management Engagement and support of participants and providers Pilot program featuring wireless monitoring of blood pressure at home
Target population	Medicaid, Medicare, and dually eligible beneficiaries who are at high risk for a major adverse cardiovascular event—defined as a heart attack, stroke, or sudden death due to cardiovascular complications—and are either not on the evidence-based medication bundle or are on some evidence-based medications but whose blood pressure was not adequately controlled (suggesting the need for review of dosing or medication adherence)
Theory of change/ theory of action	Providing participants with a health coach and appropriate evidence-based medication will reduce the incidence of cardiovascular events, improve survival rates, and reduce overall health care costs.
Payment model	New fee-for-service (FFS) payment and value-based payments
Award amount	$5,820,416

[3] The 10 medical groups were (1) Sharp Rees Stealy, (2) the Scripps Foundation, (3) the University of California at San Diego Family Medicine Group, (4) Vista Community Clinic, (5) Neighborhood Healthcare, (6) Arch Health Partners, (7) San Ysidro Health Center, (8) the North Coast Family Medical Group, (9) the University of California at San Diego Internal Medicine, and (10) North County Health Services.

Table I.1 *(continued)*

Program characteristic	Description
Launch date[a]	January 19, 2015
Program setting	Provider-based (primary care physicians)
Market area	Urban, suburban
Market location	San Diego County
Target outcomes	• Lower participants' mortality rates • Enhance participants' experience with physicians and staff in physicians' office (data come from the Clinician and Group Consumer Assessment of Healthcare Providers and Systems survey [CG-CAHPS]) • Decrease total Medicare Part A and B cost calculations • Decrease rate of emergency department (ED) visits • Decrease incidence of major adverse cardiac events • Increase percentage of participants adhering to medications

[a]After the initial planning period, the awardee's program became operational as of this date.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we conclude that University of California at San Diego was successful at implementing its program by the end of the three-year cooperative agreement. We based this conclusion on six factors.

First, University of California at San Diego served 3,731 participants—97.8 percent of its final enrollment projection—by the end of the three-year cooperative agreement. There were some challenges to recruiting and enrolling participants, and some of the partner organizations were more successful than others in meeting their individual enrollment goals. However, the program as a whole enrolled enough people early enough to make it likely that their exposure to the intervention would reveal possible effects on outcomes by the time the program ended.

Second, the awardee delivered intervention services that were consistent with its planned approach. The duration and quality of the core components—to give participants supportive, ongoing health coaching and to introduce and provide education on evidence-based medication bundles to participants and providers—remained largely the same throughout. The leadership team gave each organization the flexibility to determine the precise mode, frequency, and intensity of the delivered services, any of which varied according to the partner organizations' internal dynamics and issues.

Third, efforts by the awardee and its partner organizations to recruit and hire qualified staff were, for the most part, successful (except for some delays in hiring additional staff later on), as was their ability to retain most of their staff throughout the cooperative agreement. Toward the end of the cooperative agreement, there were some challenges in retaining staff who were suffering from burnout and worried about their job security and the program's funding going forward. The awardee also provided effective, comprehensive training in its service delivery protocol, although several frontline staff noted that in the program's final year, more training on ways to retain participants would have been useful.

Fourth, HSF-Z leaders were able to successfully engage providers at the partner organizations to support the delivery of intervention services in a timely manner and in a meaningful, sustained way.

Fifth, once the participants were successfully recruited and enrolled in the program, each partner organization engaged them in a timely manner (for example, through regularly scheduled appointments) and in a meaningful way (by individualizing services to meet their needs) and retained over half of them for the full period of enrollment. Toward the end of the program's implementation, program staff said it was more difficult to keep some participants engaged because they were feeling better. The awardee reported retaining 2,719 of the 3,731 participants through the end of the program.

Finally, both leaders and other staff said the program had a positive impact on the quality of care and the services they provided to participants, as well as on participants' quality of life and satisfaction with their care.

Impact evaluation. Mathematica will not do an impact evaluation for the HSF-Z program because the lack of clinical values in claims data and inconsistent recruiting practices across participating medical groups precluded construction of a good comparison group.

Payment model. HSF-Z leaders proposed a per-beneficiary annual payment as a model for their cardiac health intervention. The payment model would include adults with hypertension, hyperlipidemia, and/or diabetes who were either newly diagnosed or had difficulty managing their condition(s). The covered services would include an assessment of social risk factors, connections to community resources, an assigned community health worker, and a health coach for people who needed more intensive support. At the end of the third program year, HSF-Z was in negotiations with a health plan that included Medicaid, Medicare, and commercial enrollees.

Sustainability plans. A downsized HSF-Z program will continue after the cooperative agreement ends. The awardee will no longer fund health coaching or case management, but some of the participating medical groups will continue to support these activities. As it proceeds with negotiations for its payment model, the awardee plans to downsize the case management function performed by its health coaches. However, the implementing sites expect to sustain the program to varying degrees and through different mechanisms. HSF-Z leaders have presented business cases to their partners to help them with sustainment plans, but are not directly involved in helping sites tailor those plans.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded around the start of the third program year and achieved a response rate of 95 percent from the sample size of 22. We did not weight the survey samples to adjust for nonrespondents. For survey items that fewer than 11 people responded to, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that University of California at San Diego successfully implemented its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

University of California at San Diego's leadership team originally partnered with eight San Diego-area medical groups to implement HSF-Z. It added two more organizations during the second year of the program, for a total of 10 partnerships. Although the leadership team did provide support and guidance, it gave each partner organization the flexibility to determine its own best approach to identifying, recruiting, and enrolling participants.

The partner organizations took different approaches to identify eligible participants from their existing patient population, and the organizations modified their strategies throughout the implementation period. For instance, one organization optimized its patient registry so that program staff could search for patients with cardiovascular risk factors who are due for their

annual physical exam. Another organization that was initially offering the program only to patients with diagnoses of high blood pressure, cholesterol, or diabetes began calculating a 10-year cardiovascular event risk score[4] for patients, and used this information to identify high-risk patients who might benefit from the program.

Recruitment and enrollment practices varied by organization, and the different organizations modified their approaches throughout the implementation period as well. Some organizations recruited and enrolled all participants in person, either when they came into the clinic for other types of appointments or by scheduling them to come in solely for this purpose. Some organizations found "warm hand-offs" to be an effective approach to recruiting and enrolling participants. The provider introduced the HSF-Z program to the patient during a regular appointment, and then referred the patient to the health coach to be enrolled directly after the appointment. Other organizations enrolled participants by mailing or emailing the program forms to them, and had them return the forms by mail or email. Many of the organizations originally planned to identify, recruit, and enroll participants from only one or two of their clinics. Most of them revised their strategy, however, and opened the program up to all of their clinics.

Over the course of the cooperative agreement, University of California at San Diego extended the enrollment deadline from May 2016 to August 2016, and then extended it again to February 2017.[5] This gave more time to some of the partner organizations that were still experiencing difficulties in enrollment. University of California at San Diego also provided for flexibility in the inclusion and exclusion criteria on the medication bundle requirements, allowing, for example, certain patients on other evidence-based medicine to qualify for the program if the patient was allergic to or otherwise unable to take the original bundle for some reason.

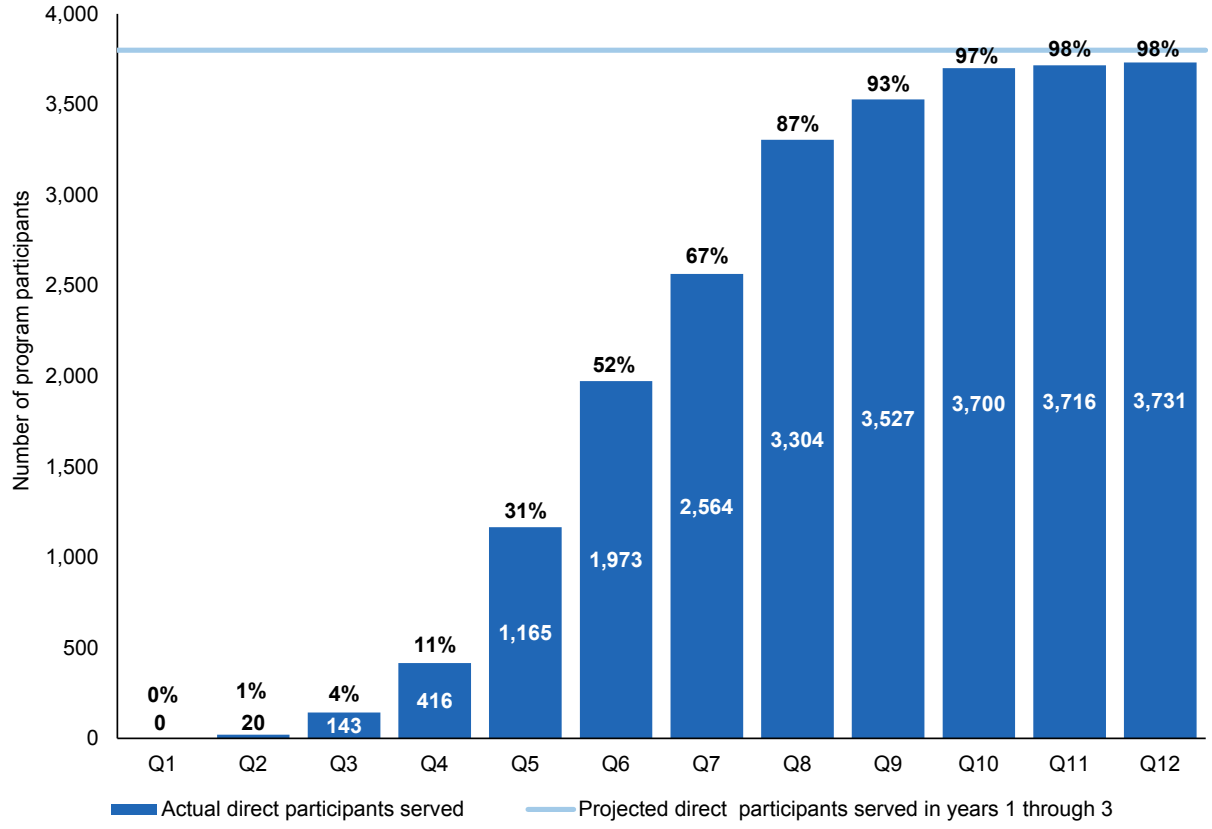
b. Evidence of enrollment effectiveness

University of California at San Diego reported that 3,731 participants had spoken at least once with health coach from January 2015 (when it launched its program) through August 2017. This represents about 98 percent of its revised target of 3,800 projected participants over three years (Figure II.1). However, the group of 3,731 participants does not include people who signed up for the program and did not return for services. University of California at San Diego revised its original three-year target of 4,008 participants to 3,600 in the beginning of the second program year, and then updated the target to 3,800 in the beginning of its third program year. Program leaders and frontline staff implemented new strategies throughout the duration of the program in order to reach this goal, including adding partner organizations and clarifying eligibility criteria. Partner organizations also added sites within each medical group to increase the eligible population. They revised their individual strategies for identifying their eligible patients and for recruiting and enrolling them into the program. The strategies of University of California at San Diego and its partners resulted in an effective enrollment effort.

[4] The 10-year cardiovascular event risk score estimates a patient's risk of coronary death, heart attack, or stroke within the next 10 years. The score is based on an algorithm published in the 2013 American College of Cardiology/American Heart Association Guideline on the Assessment of Cardiovascular Risk.

[5] However, five participants were enrolled in the quarter immediately following February 2017.

Figure II.1. Projected versus actual cumulative direct participants ever served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. University of California at San Diego revised its original three-year target of 4,008 participants to 3,600 in the beginning of the second program year, and then updated the target to 3,800 in the beginning of its third program year.

c. Barriers and facilitators associated with enrollment effectiveness

University of California at San Diego's progress in meeting its three-year enrollment goals was influenced by several factors. Initially, the HSF-Z program experienced significant delays in receiving approval from the institutional review board (IRB), which pushed the launch date of the program back and delayed the start of enrollment at each organization. Once they started, recruitment and enrollment also proved to be more challenging than the HSF-Z leaders and frontline staff had envisioned.

The program team did not anticipate that so many eligible patients would decline to participate in HSF-Z because they did not see it as necessary or important. Over half of the staff surveyed also noted that language was a barrier for potential participants. However, HSF-Z leaders reported that changing recruitment and enrollment strategies—such as replacing health coaches' initial approach of cold-calling eligible participants, developing more culturally

appropriate messaging, and others—helped organizations overcome this challenge. In particular, leaders and staff noted that an important element of success was having physicians be aware of the program and supportive of it, including having an active physician champion at each organization. For instance, in some organizations providers got more involved in recruitment by introducing the program and giving potential participants warm hand-offs to health coaches.

2. Delivery of program services

a. Description of and changes to service delivery model

Frontline staff at each organization identified participants at high risk for major adverse cardiovascular events. Once participants were enrolled, health coaches worked with physicians to ensure the participants were put on appropriate, evidence-based medication bundles for hypertension, diabetes, and other conditions that put people at high risk for cardiovascular disease. The health coaches met monthly with participants, also educating them about good health habits and helping them set and achieve their health goals—both to improve compliance with the drug regimens and to encourage lifestyle changes.

Some partner organizations also introduced a wireless blood pressure monitoring program in Year 2. The blood pressure cuffs had a memory storage function that allowed participants to access readings. The blood pressure data were transmitted to a secure server, which health coaches could access and use to generate reports. University of California at San Diego reported that the program was successful in engaging participants, transmitting timely data to care teams, and improving teams' ability to respond quickly with treatment changes. The HSF-Z program has also been supplemented by a community-wide effort to educate patients on the risks for cardiovascular disease (the Be There program) and a long-standing effort to educate physicians as well.

Finally, during the third year of the program, University of California at San Diego leaders changed the requirements for participant engagement. In response to lowered participant engagement over the course of the program, leaders looked for ways to decrease the amount of time participants needed to devote to the program. Health coaches were allowed to retain participants in the program, even if they did not meet monthly with them, giving some participants a break from monthly meetings. Some of the participating organizations also worked to roll out a pilot text messaging program. The goals of the pilot program were to (1) allow coaches to obtain information by text from participants in between meetings so staff could prepare for their coaching visits or calls and (2) allow participants to ask questions in between monthly meetings. As of the program's end in August 2017, one of the partner organizations had 158 participants using the texting program. Some participants liked the flexibility in timing that texting allows, and several who were unwilling to continue the monthly scheduled phone calls after meeting their health goals were able to stay enrolled through the texting program.

Several participating organizations already had some of the program components in place before the launch of the program. For example, some participating organizations were already employing health coaches as coaches for other programs, or the health coaches already had other roles and responsibilities in the organizations before they joined the program. Some organizations already had their own home blood pressure monitoring program in place (although these programs typically were not wireless). In addition, most organizations already had in place

other initiatives or internal programs to improve the quality of care for many patients in the target population for HSF-Z, such as those with diabetes.

b. Evidence of service delivery effectiveness

University of California at San Diego was successful at implementing its program in a timely manner. University of California at San Diego leaders allowed for flexibility among the partnering organizations in how services were delivered (for example, by phone or in-person), however, the primary components of having health coaches provide care management and of encouraging and supporting participants' adherence to evidence-based medicine bundles were consistently and successfully implemented. Health coaches engaged their participants in a meaningful and timely manner, though the program leaders did note some challenges to retaining participants after they had been enrolled and met their goals. University of California at San Diego was able to recruit, hire, and retain staff successfully, and the rapport between health coaches and participants was seen as pivotal to their success. Staff also noted that University of California at San Diego provided effective and comprehensive training and kept providers and provider organizations engaged throughout the program. Details follow.

Delivery of intervention services. Based on the evidence available to us, we conclude the partnering organizations successfully delivered the core components of the program: providing care management through health coaches and encouraging and supporting participants' adherence to evidence-based medicine bundles**.**

The results of the qualitative interviews indicate that the intended quality and duration of services stayed consistent throughout the program implementation period. The frequency, intensity, and mode of services varied by the partnering organization, however. Once they completed the hour-long intake session (generally, in person but occasionally over the phone), health coaches generally spoke with participants for about 45 minutes every month, though the frequency and intensity of these meetings varied with the needs of the participant. For example, one health coach noted that appreciated the flexibility to be able to sit with the patient for an hour if that was what it took for them understand the results of their health test.

After the initial session, many coaches continued talking with participants over the phone for the most part. Some health coaches worked at one of the organizations' participating sites, whereas others traveled to different locations to meet with participants. Some health coaches worked remotely and did telephone follow-ups, but one health coach made in-person home visits to some participants with higher social risk factors and/or fewer community resources.

Secondary services of HSF-Z included the pilot program of wireless transmission of blood pressure readings taken at home, as well as education for patients at a community level, and provider education. These secondary services were also successful.[6] During the third year, the success of the pilot texting program varied based on the patient population. Several sites reduced contact with participants who had met their health goals and begun to lose interest, and texting was seen by some staff as a way to decrease the burden on participants (as well as the health coaches' own workload).

Staffing and training. Participating organizations were able to hire the health coaches who were an essential component of the intervention, working with physicians to implement the program. Program leaders said that several organizations already had health coaches on staff at the start of the program, whereas others had to hire coaches as part of their agreement to participate in the program. Most of the organizations employed the health coaches full-time. At other organizations, the health coaches worked part-time on HSF-Z and had other roles and responsibilities. In interviews at partnering organizations, some organizations attributed their success in enrollment to hiring health coaches who exclusively worked on HSF-Z, noting that it was easy for other staff to be pulled away to work on other projects. Program leaders as well as participating organizations also noted that the health coaches' background and experience varied greatly by organization as well (some were registered nurses or pharmacists, for example), but most had limited clinical training and had been health educators or navigators in earlier positions.

> "There was a lot of flexibility in terms like what the definition of a health coach was … but everybody's populations are so different, [the program leaders] definitely did it that way because they had to."
>
> —Health coach

The University of California at San Diego gave the health coaches and partnering organizations training and technical assistance, including convening two in-person training sessions a year for the health coaches. Although the majority of staff surveyed by Mathematica said the training was helpful to their job, and almost half of staff surveyed thought that more training would have been helpful. Some suggested training on burnout, or on how to keep participants engaged in the final program year.

> "A very major determinant was having a physician champion. In those places where we had strong physician champions, the program tended to be very successful. In those settings where we had to reach out to somebody, secondarily, it was a bit more challenging. [In] dealing with different patient populations and different health care settings of different medical organizations … the one thing that seems to go above and beyond all those variables was [having] a really strong physician champion. Regardless of the setting, you were a lot more likely to have success."
>
> —UCSD leader

The awardee also held monthly meetings with the health coaches. The frontline staff reported that these in-person sessions and the monthly meetings among the different health coaches were useful. These meetings were an

[6]The University of Best Practices (UBP) is an effort to get medical directors and providers involved in voluntary collaborations to improve clinical practices, including pharmaceutical therapies used in primary intervention. Be There San Diego is a countywide initiative to educate the San Diego population on how to prevent heart attacks and strokes and thus "be there" for their loved ones. Both the UBP and the Be There initiative predated HCIA R2; however, HSF-Z leaders are using them to provide further education on the program.

opportunity for health coaches to learn from each other and troubleshoot common challenges together.

Recruitment and engagement of providers. University of California at San Diego leaders said the program was successful in recruiting provider organizations to implement HSF-Z. From the beginning of the implementation period, University of California at San Diego's leadership team said it had partnerships with eight organizations and successfully added two provider organizations during the second year of the program. Both the leaders and frontline staff we interviewed told us they were also successful in engaging providers and provider organizations. Most of the partnership organizations had "provider champions," and frontline and leadership staff both said these provider champions were critical to the success of implementation.

Interviews with frontline staff and leaders and the results of the staff survey indicate that partner organizations or sites without a strong physician champion noticed a difference in provider engagement. More than half of the staff surveyed believed providers might be reluctant to engage because they expected the providers would consider their current model of care to be sufficient. To help encourage providers to engage with the program, the University of California at San Diego's program director visited sites to discuss the program with providers as a form of peer-to-peer encouragement. Some frontline staff we interviewed said these visits helped increase awareness and willingness to participate more so than they were able to do on their own.

Engagement of program participants. The awardee had some challenges with recruitment and enrollment. There were not as many eligible patients as expected who were interested in joining the program, and there were also cultural and language barriers for eligible patients. University of California also found that of the 4,158 people that initially signed up for the program, about 400 never engaged with any of the services. In response, organizations revised their recruitment strategies to include more upfront education on the benefit of joining the study as well as more culturally sensitive materials and support. Over the course of the program, the health coaches were able to enroll and then engage participants in a meaningful and timely manner, and the relationships built between the health coaches and participants were considered integral to the success of the program.

During the third year of the cooperative agreement, health coaches struggled to retain participants, especially those who had met their goals. These participants did not see the benefit of continuing in the program. HSF-Z leaders and staff said the health coaches built a rapport with participants, and this trusting relationship both helped the program succeed and kept many participants from disenrolling. Staff also noted they used other incentives such as gift cards to help keep some participants engaged and encourage them to provide survey data, but this was only effective for the participants the health coaches were able to reach. Over a quarter of the participants who had engaged with the program ended up disenrolling (27 percent).

The majority of staff who were surveyed thought participants considered the program to be too much of a time commitment. To reduce the burden on participants, University of California at San Diego developed a texting program as a way for participants to quickly and easily update their coaches on their progress. The text program successfully engaged participants at some sites, but at other sites participants reportedly did not use the program. Program leaders then loosened

restrictions on enrollment, which allowed health coaches to keep participants enrolled even if they were unable to connect each month.

Awardee leaders noted that the trusted relationships the health coaches had established with participants helped keep disenrollment levels low. Overall, health coaches were able to retain and engage the majority of participants throughout the duration of the program. Of the 3,731 participants receiving a service, 2,719 were retained through the end of the program.[7]

c. Barriers and facilitators associated with service delivery effectiveness

The University of California at San Diego and its implementing organizations identified the following barriers to meeting their enrollment goals. First, the IRB approval process delayed the launch date and consequently diminished the organizations' ability to enroll participants as early as they hoped to. Second, more eligible patients than expected declined to participate. Partner organizations reported that this happened because some patients did not think the program would be helpful to them, whereas others had concerns about the statin medications that are a required part of the medication bundle. In some cases, reluctant patients were motivated to enroll and engage after their provider endorsed the program and explained the benefits of participation. University of California at San Diego leaders and frontline staff believed increased physician involvement facilitated enrollment. Staff also thought the strong rapport between health coaches and participants was pivotal to successful implementation.

> "A big takeaway is the unique relationship of [the] health coach. They bring so many strengths … the relationships that our health coaches really had on an emotional level and trusting level with our patients initially really resonated all the way throughout the program. It was that relationship that they had [with the participants] that prevented them from ending [their participation]."
>
> —Health coach

Several organizations as well as the program's leaders noted the challenge of keeping participants engaged in the program during the final year of the program. It was also difficult to retain participants who thought they were doing well and no longer needed or wanted the services provided by health coaches. However, staff noted that the trusting relationships formed between health coaches and participants kept the disenrollment rate low. In addition, during the latter part of the implementation, most (90 percent) of the enrollees were medically adherent, and the health coaches put more emphasis on secondary goals of self-management and other lifestyle-focused goals such as diet and exercise to keep participants engaged. Some frontline staff also noted that they provided incentives in the form of gift cards.

C. Assessment of perceived program effects on the delivery of care and outcomes

Mathematica's surveys and interviews with staff implementing the program, as well as HSF-Z's self-monitoring and measuring data, indicate that HSF-Z had a positive effect on

[7]The implementation and monitoring contractor enrollment data and awardee self-reported enrollment data cite a final enrollment of 4,158 people, and 3,731 participants served, and the awardee reported in interviews that 2,719 participants were retained through the end of the program. However, the program's enrollment data show a final enrollment figure of 2,718 participants. According to these data, 4,158 participants enrolled, 24 declined, 508 withdrew, and 908 were removed.

participants. The majority of staff surveyed thought the program had improved their organization's ability to provide care to participants on nearly all of the patient-focused measures, such as helping participants achieve their health goals; providing care that is responsive to patient needs, preferences, and values; and improving participants' overall quality of life. In addition to their own experiences with participants, staff cited several other sources of information on which they based their perceptions of HSF-Z's success. Early evidence collected by sites and HSF-Z leaders suggested improved blood pressure and diabetes control, better adherence to medication regimens, and more prescribing of evidence-based medications.

Results of participant surveys fielded by HSF-Z staff suggested that as a result of HSF-Z, program participants felt more connected to their clinics, better able to navigate the health system, and more confident about advocating for their care. Program staff interviewed and surveyed by Mathematica also thought the program helped participants develop a better understanding of which issues they should be seen for at the clinic and which issues could be managed at home, reducing the number of ED visits and unnecessary visits to a provider's office and consequently lowering costs. Most staff surveyed also thought that HSF-Z allowed their organizations to run more efficiently and improve access to care, as well as provide care and services fairly to all patients. At one site, an interviewee noted that the program reporting revealed gaps in data and care at the organization. As a result, the site instituted more consistent procedures for scheduling cholesterol labs, doing blood pressure rechecks, and calculating total cholesterol for a patient.

In interviews and in the staff survey, program staff also said the program could change in a few key ways. Specifically, most staff thought the health coaches' participant load was too high, and that a smaller number of participants would be easier for health coaches to regularly and effectively engage. Staff also said the length of participation in the program could be defined instead of indefinite, and some suggested a year as an appropriate amount of time. Some health coaches, in interviews and the staff survey, conveyed their sense that participants did not want to be bothered once they met their health goals. These suggestions might improve, or at least not negatively affect, program outcomes, according to participants.

Although more than half of the staff in our staff survey said the program increased their feelings of burnout at work, and half noted a high administrative burden, the majority of staff said that HSF-Z was worth the effort overall.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

Our findings indicate that University of California at San Diego was able to successfully implement its program. We assume, as did the awardee's theory of change/theory of action, that giving participants a health coach and appropriate evidence-based medication will reduce the incidence of cardiovascular events and ED visits, improve survival rates, and lower health care costs. The awardee was able to enroll 4,148 people and served 3,731 participants, but was only able to retain 2,719 through the final quarter. About 400 people signed up, but received no

services, and over 1,000 participants received at least one service from a health coach, but not to the degree the awardee intended. The total number of participants served, 3,731, includes participants who only received an intake appointment, as well as participants who chose to disenroll once reaching their health goal. However, we can only say with with certainty that 2,719 participants received the full amount of intended services. We therefore can reasonably conclude that the HSF-Z program has made progress toward achieving those outcomes, but effects may only be seen among the 2,719 participants retained.

Despite delays in obtaining IRB approval, University of California at San Diego was able to come close to its enrollment goals, deliver services as intended, and deliver services to all participants for several months. Health coaches monitored participants' health monthly, which allowed the awardee to see progress on a relatively regular basis. Although University of California at San Diego relaxed the monthly meeting requirement in the third program year, health coaches indicated that most participants met their health goals before choosing to meet less than once a month with their health coaches.

Participant outcomes will likely vary slightly by site, especially since sites varied widely in the organization resources they could devote to the program. Some health coaches thought the program was more difficult to implement in smaller sites. For example, some sites already had similar and complementary programs that served an overlapping group of patients, and were able to use resources from those programs to supplement implementation of HSF-Z. This may also be a factor in whether a site decided to sustain the program, and these sites may be able to provide more evidence of program effectiveness down the line.

Data collected by the awardee as part of its self-monitoring measures, along with anecdotal evidence, suggest an uptick in interim effects, such as better adherence to evidence-based medication and healthier blood pressure levels for participants with hypertension. Health coaches shared a variety of stories with us about participants who lost weight, lowered their blood pressure, and made positive overall lifestyle changes, and some coaches reported that their sites' tracking of health indicators shows lowered risk factors for participants.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

In this chapter, we present our assessment of the feasibility of conducting a rigorous impact evaluation. We also report baseline characteristics of the Medicare FFS treatment group, which we identified using the awardee's enrollment finder file. The University of California at San Diego ended enrollment in February 2017, so the baseline characteristics are reported for their final Medicare FFS study population.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data. The University of California at San Diego program ended August 30, 2017.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: University of California at San Diego

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure	1,176[a]
Projected Medicaid population with 6 months of program exposure	214[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	2,816
Likelihood of all-cause hospitalizations	4,154
MDE sample size requirement to detect 20% effect	
Total expenditures	704
Likelihood of all-cause hospitalizations	1,039
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Questionable, effects may not be observed in follow-up period
Claims sufficient to identify treatment and comparable comparison group?	Questionable, no testing yet to determine strength of intent-to-treat framework
Likelihood of solid comparison group	Serious concern; we may be not able to identify a strong comparison group
Do claims identify the primary expected effects?	Some effects observed in claims data, but important effects likely missing

Table III.1 *(continued)*

Evaluability domain	Response
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Lack of strong comparison group
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We do not anticipate being able to construct a valid comparison group for the awardee. Therefore, we are unable to conduct a rigorous impact evaluation. The treatment beneficiaries are identified by using clinical data that are not available in claims data. We have received clinical information that is provided by clinicians and used to generate baseline and follow-up cardiovascular risk scores for stroke and acute myocardial infarction. We evaluated the completeness of these data to determine if we could use these data in an analysis of whether the awardee's program resulted in a change over time in cardiovascular risk. Less than one-quarter of the beneficiaries enrolled in the program had sufficiently complete clinical data to recalculate atherosclerotic cardiovascular disease (ASCVD) risk. Since these patients are unlikely to be representative of all enrolled patients, results from such an analysis are likely to be subject to bias. We therefore conclude that the administrative data cannot be used to generate a reliable estimate of whether and the extent to which the program reduced ASCVD risk. We will report on the experiences of staff and participants, based on our surveys.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents baseline characteristics of the treatment group, which we measured during the 12 months before each beneficiary's enrollment date for the HSF-Z program. The treatment group consists of eligible patients at high risk of heart attack or stroke who were recruited by participating physician practices. To support an intent-to-treat evaluation, the treatment group would also include eligible patients who were recruited but refused to participate in the program. However, with the exception of a small number of cases, the University of California at San Diego was unable to obtain the identities of these patients. The treatment group does include those who withdrew from the program or were removed (mostly for noncompliance). These patients comprise 31 percent of the treatment group.

Participating physician groups began enrolling Medicare and Medicaid beneficiaries in the HSF-Z program on January 19, 2015. As of the end of February 2017, the program had 4,158 participants, 2,863 of whom were linked to administrative records.[8] Among all participants,

[8] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

2,651 were FFS Medicare beneficiaries, and 835 participants were Medicaid beneficiaries. The remaining participants included 1,297 individuals whom we were unable to link with administrative data.[9] In presenting the baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they must have been enrolled in the awardee's program on or before February 28, 2017, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. After we excluded Medicare beneficiaries who did not meet the above criteria, a total of 1,130 participants were included in the analysis of baseline characteristics for this report. We were unable to report baseline characteristics of Medicaid beneficiaries in the study population because we lacked current claims data on them.

These Medicare FFS beneficiaries look generally similar to the Medicare FFS population nationwide. Their characteristics are shown in Table III.2. Beneficiaries younger than 65 account for 9 percent of program participants, whereas 65 percent are ages 65 to 74. Only 2 percent are older than 85. The percentage of females is slightly lower than the national average for all Medicare FFS beneficiaries: 52 percent versus 54 percent. The Medicare FFS beneficiaries in the program are also less likely than the general population of Medicare FFS beneficiaries to have originally been eligible for Medicare because of a disability (17 percent versus 24 percent). They are more likely to have been eligible because of old age and survivor's insurance (83 percent versus 76 percent). Less than 1 percent were eligible because they had end-stage renal disease (ESRD). About one in five (19 percent) are dually eligible for Medicare and Medicaid. The average hierarchical condition categories (HCC) risk score of 1.15 is moderately higher than that of the overall Medicare FFS population. The HSF-Z program appears to have a somewhat greater number of predicted higher-cost patients relative to the national Medicare FFS beneficiary population, as indicated by the interquartile range of the HCC score (0.47 to 1.4 versus 0.49 to 1.25).

Table III.3 shows baseline utilization and expenditure data for a common set of measures, including the Center for Medicare & Medicaid Innovation's (CMMI) four core measures. We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $867, which was somewhat higher than the national average in 2014 ($792). Among HSF-Z participants, the average PBPM Medicare payments for physician services ($286) and for acute inpatient care ($287) were the largest drivers of the total cost of care in the baseline year. The rate of acute hospitalization among HSF-Z participants (247 per 1,000 participants) remained below the 2014 national average (274 per 1,000 beneficiaries). The rate of ED visits not resulting in a hospitalization or in an observation stay is somewhat lower than the national average (638 per 1,000 participants versus 652 per 1,000 beneficiaries).

[9] Some of the larger participating health systems failed to report Social Security numbers, HICs, or other insurance identifying numbers, to the extent that this information was lacking for a majority of participants. For most of these enrollees, we were able to link to claims by using name, date of birth, gender, and zip code.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the University of California at San Diego's program through February 28, 2017

	All participants (n = 1,130)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	106	9
65 to 74	736	65
75 to 84	266	24
85 and older	22	2
Gender		
Female	585	52
Male	545	48
Race		
White	839	74
Black	84	7
American Indian, Alaska Native, Asian/Pacific Island American, or other	101	9
Hispanic	86	8
Original reason for Medicare eligibility		
Old age and survivor's insurance	933	83
Disability insurance benefits	188	17
End-stage renal disease (ESRD)[a]	9	0.8
Hospice[b]	1	0.09
Medicare/Medicaid dual status, percentage dual[b]	216	19
HCC score[c]		Statistic
Mean		1.15
25th percentile		0.47
Median		0.81
75th percentile		1.4

Source: Mathematica analysis of information from awardee's finder file as of February 28, 2017, and Medicare claims and enrollment data as of August 30, 2017.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the beneficiary consented to participate in the program. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the University of California at San Diego's program through February 28, 2017

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	1,130	1,047	1,095	1,129	1,130
Average Medicare expenditures PBPM[a]					
Total	867 (50)	845 (82)	868 (76)	886 (76)	866 (73)
Acute inpatient	287 (26)	308 (57)	262 (45)	308 (46)	272 (50)
Inpatient other[b]	14 (5)	23 (16)	4 (4)	15 (8)	14 (10)
Outpatient[c]	173 (18)	167 (18)	169 (23)	192 (29)	162 (16)
Physician services	286 (13)	287 (20)	277 (17)	270 (13)	308 (16)
Home health	44 (8)	25 (7)	55 (12)	47 (8)	49 (11)
Skilled nursing facility	39 (9)	24 (10)	67 (22)	37 (13)	26 (12)
Hospice	2 (2)	0 (0)	0 (0)	2 (2)	5 (5)
Durable medical equipment	23 (4)	13 (2)	34 (11)	15 (3)	29 (7)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	247 (20)	226 (34)	244 (37)	289 (37)	227 (31)
Outpatient ED visits[e]	638 (131)	545 (107)	631 (159)	613 (150)	755 (149)
Primary care visits in any setting	5,587 (159)	5,260 (235)	5,312 (221)	5,470 (228)	6,246 (206)
Primary care visits in ambulatory settings	4,998 (131)	4,656 (185)	4,779 (181)	4,846 (177)	5,651 (177)
Specialist visits in any setting	9,966 (330)	9,982 (419)	9,731 (405)	9,865 (424)	10,246 (418)
Specialist visits in ambulatory settings	8,746 (284)	8,728 (354)	8,627 (344)	8,577 (347)	9,016 (344)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	17 (1)	5 (1)	6 (1)	6 (1)	5 (1)
Percentage with an outpatient ED visit[e]	26 (1)	8 (1)	9 (1)	8 (1)	11 (1)
Percentage with a 30-day readmission among all discharges	14 (2)	14 (5)	14 (4)	12 (4)	13 (4)
Percentage of participants with a readmission among all participants	2 (<0.5)	1 (<0.5)	1 (<0.5)	1 (<0.5)	1 (<0.5)

Source: Mathematica analysis of information from awardee's finder file as of February 28, 2017, and Medicare claims and enrollment data as of August 30, 2017.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

These results suggest that the University of California at San Diego did not recruit a patient population with particularly high costs of care. However, the HSF-Z eligibility criteria focus on identifying people at risk of high health care costs in the future, not on people whose current utilization is high.

The likelihood of a 30-day all-cause hospital readmission among those hospitalized during the baseline year was 14 percent, below the national average for Medicare FFS beneficiaries (18 percent). Only a small share of both primary care and specialist visits (11 percent and 12 percent, respectively) were provided outside of ambulatory care settings, reflecting relatively low hospitalization rates in the program population. Unlike results in previous reports on the University of California at San Diego's patients, where average PBPM expenditures were found

to fall during the course of the baseline year, total and component PBPM expenditures are relatively stable.

To better characterize the cardiovascular risk among beneficiaries in the treatment group, we measured (for the calendar year preceding the year of their enrollment) the prevalence of chronic conditions that are associated with an elevated risk for major cardiovascular events (Table III.4). We used the Chronic Care Warehouse (CCW) condition flags to identify beneficiaries who had ever been diagnosed with each condition. Compared with all Medicare FFS beneficiaries in 2015, beneficiaries in the treatment group were more likely to have diabetes (39 percent versus 27 percent), hyperlipidemia (51 percent versus 45 percent), and hypertension (64 percent versus 55 percent). However, the incidence of ischemic heart disease, heart failure, and prior stroke or transient ischemic attack was similar to national averages.

Table III.4. Prevalence of related chronic conditions among Medicare FFS beneficiaries in the treatment group relative to all Medicare FFS beneficiaries

Percentage of beneficiaries with	Treatment group	All Medicare FFS beneficiaries
Acute myocardial infarction	1	1
Atrial fibrillation	6	8
Diabetes	39	27
Heart failure (congestive heart failure)	13	13
Hyperlipidemia	51	45
Hypertension	64	55
Ischemic heart disease	26	26
Stroke or transient ischemic attack	4	4
Number of observations	1,130	

Source: Mathematica analysis of information from awardee's finder file as of February 28, 2017, and Medicare claims and enrollment data as of August 30, 2017. Prevalence rates for all Medicare beneficiaries are from the *Chronic Conditions Among Medicare Beneficiaries, 2015*, available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Chartbook_Charts.html.

Notes: The data in this table are based on CCW "ever had" condition flags during the calendar year preceding the year of patients' enrollment in HSF-Z and may not precisely replicate the clinical definitions used by UCSD's partner practices to identify patients with specific chronic conditions.

CCW = Chronic Conditions Warehouse; FFS = fee-for-service.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

At the end of the third year of the cooperative agreement, University of California at San Diego had developed a payment model to support its cardiac health intervention. The payment model was based on a per-beneficiary annual payment. The awardee was in preliminary negotiations with a health plan that had Medicaid, Medicare, and commercial lines of business, and reported that other plans had expressed interest in the model pending more analyses of program outcomes.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

The awardee proposed a per-beneficiary annual payment as a model for its cardiac health intervention. The payment model would include adults with hypertension, hyperlipidemia, and/or diabetes who were either newly diagnosed or had difficulty managing their condition(s). The covered services would include an assessment of social risk factors, connections to community resources, an assigned community health worker, and a health coach for individuals requiring more intensive support.

Based on an analysis of potential savings from reduced cardiovascular events using the Archimedes modeling software, the awardee estimated annual payments per beneficiary from the payer to the program of $620 for Medicare Advantage plans ($51.67 PBPM), $559 for commercial plans ($46.58 PBPM), and $182 for Medicaid plans ($15.17 PBPM). These estimates also incorporated a discount for the average proportion of beneficiaries that are enrolled in one year but not in the next, for whom plans would not realize reduced costs due to fewer events. The awardee did not report any plans to adjust these rates based on beneficiary characteristics, medical complexity, or measures of quality or spending.

C. Status of the payment model

At the end of the third year of the cooperative agreement, the awardee's leaders reported that they were in preliminary negotiations with a health plan that included Medicaid, Medicare, and commercial enrollees. They also noted that they continued to engage with a few other plans that had expressed interest in the program, but wanted to see more concrete evidence of results and return on investment. The awardee leaders planned to continue marketing the program and payment model to organizations (including managed care plans and accountable care entities) that might incur health care costs for adults with the targeted conditions.

D. Factors associated with the development of the payment model

The leaders of the HSF-Z program said three things facilitated their development of a payment model. First, awardee leaders described how their experience using data to guide the

program also helped them perform the analyses necessary to develop the payment model. Second, they noted that a community health benefit requirement from a recent health plan merger might have encouraged the plan's interest in the program. And third, the awardee leaders said they had been watching the experiences of a Round 1 HCIA awardee that experienced slow, deliberate growth after its cooperative agreement. This informed University of California at San Diego's payment model planning.

The awardee leaders also described three major barriers to their development of a payment model. First, the awardee leaders noted that a primary challenge in engaging payers was the program's focus on averting cardiovascular events that might not happen for several years; the health plans were more focused on costs over a much shorter time frame. This challenge was exacerbated by the fact that many beneficiaries churn between plans from one year to the next, making health plans reluctant to invest in an intervention they might not benefit directly from. Second, the awardee leaders noted that their proposed payment model was for a centralized program that would still require significant startup funding after the cooperative agreement, and the proposed payment model would cover operating costs, but not the investments needed upfront. The awardee reported that it plans to seek other investors to cover the extra costs. Third, the awardee leaders noted that accountable care entities affiliated with other health systems were likely reluctant to engage with them because the University of California at San Diego health system was viewed as a competitor. To help address this, the awardee has established an independent 501(c)(3) organization for the program.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The University of California at San Diego was planning to sustain a somewhat downsized HSF-Z program after the cooperative agreement ended. As it proceeded with negotiations for its payment model, the awardee planned to downsize the case management function performed by its health coaches. Sustainment plans varied across the implementing sites, however. Increasing the use of pharmacists was a goal for some, but payment issues were a potential obstacle. The awardee had early conversations about scaling the program to other populations, but did not appear to be helping other organizations replicate the program.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2 of the award, the University of California at San Diego had minimal plans for sustaining its HSF-Z program. Each implementing site was identifying sustainment strategies, including speaking with organizational leaders and incorporating the program into existing programs. The awardee proposed a population health payment model to the state Medicaid program and had plans to eventually propose something similar to other payers. At that time, the

University of California at San Diego had not scaled or replicated its program and had no plans to do so.

C. Implementing the SSR plan: progress and changes

Sustainability. In Year 3, the University of California at San Diego reported that many of the operational barriers to its HSF-Z program were sorted out during the course of the program, so sustainability efforts relied on each site's ability to justify the program's cost and secure new payment sources to maintain it. The awardee helped individual sites pull together data and other information to build a business case for sustaining and securing ongoing reimbursement to support the health coaching function. As described in Chapter IV on the payment model, the awardee was in early negotiations with one payer to provide a case rate payment for the program.

After the end of the cooperative agreement, the awardee was not supporting health coaching or case management. Reportedly, however, most of the program sites valued the health coaching function and intended to continue funding it, in many cases by wrapping the function into other case management or health coaching that is already in place for other types of patients. Because the sites had used different types of staff as health coaches—for example, nurses, pharmacists, and psychologists—it made sense for them to figure out how to sustain this function in a way that fits their existing structure.

The awardee's most successful example of a site sustaining the program is the San Ysidro Community Health Center (CHC). The CHC began employing its own health coaches in each of its sites, and it integrated health coaching into the care coordination function it implemented as part of achieving recognition as a patient centered medical home. The CHC successfully applied for a higher Medicaid prospective payment system (PPS) rate from the Health Resources and Services Administration, which provides ongoing funding to support this additional service.

One of the sites (North Coast) did not plan to sustain the program, both because of a limited budget and because the site viewed the program as a time-limited research project. North Coast decided the health coaches would no longer work proactively and intensely with the patients. However, a respondent from that site thought that encouraging the use of statins for high cholesterol had become ingrained in the care process and that medical assistants would continue to routinely check blood pressure during appointments, so some semblance of the program would be maintained.

At the sites that planned on continuing the program, health coaches were prioritizing participants who met criteria indicating they need ongoing care and supports, and those participants were being assigned to a care coordinator. The rest of the participants would return to receiving the type of care they were getting before the program. At one site (Sharp Rees), a few patients with uncontrolled issues were enrolled in the Next Generation Accountable Care Organization, and the hospital had nurses budgeted on staff to manage their care.

The awardee and some of the sites wanted to rely more on the pharmacists going forward and were focusing on how to generate funding to maintain and grow that staff. This included plans to establish criteria and processes for how pharmacists could manage medications for patients going forward, with protocols for referring patients to health coaches or nurses as needed. The University of California at San Diego Internal Medicine Department decided to

maintain the pharmacist full time, with half of the position's time spent on the in-person health coaching visits and the other half on services that provide higher reimbursement (particularly annual wellness visits for Medicare patients) to help sustain the telephone health coaching function, which is not billable. In addition, a clinical pharmacist at the University of California at San Diego created a proposal and budget for continuing the role, with help from the Be There team, on analyzing the program's early impacts and cost savings, and on developing other business strategies.

Scalability. The awardee was working on ways to provide health coaching and medication adherence functions to more participants via contracts with payers and providers. The University of California at San Diego held meetings with a health plan that seemed interested in using model for its diabetic population. The awardee reported making "good headway" with that payer and held early, informational meetings with two or three additional payers.

Replicability. The awardee did not seem to be helping other organizations replicate the program.

D. Factors associated with progress toward implementing the SSR plan

The ability to secure ongoing payment was an important part of SSR planning; the barriers and facilitators to doing that are discussed in Chapter IV on the payment model. In addition, interviews revealed several factors that are likely to affect the awardee's and the sites' ability to rely more on pharmacists going forward.

The demand for telephone interactions was relatively high for this patient population, because many of the patients have low incomes and lack transportation, or they have other problems that prevent them from seeing providers in person. However, because these interactions were typically not billable (unless the patient consented to enroll in chronic care management and paying a co-pay), the providers had to find other ways to generate revenue to support the function.

Some sites found different ways to justify to their leaders that the health coaching by pharmacists helped outcomes, and some were trying to develop collaborative practice agreements to obtain reimbursement. At the same time, California started allowing pharmacists who have physician oversight to bill at a higher level than before, and a respondent thought the University of California at San Diego would soon participate in this, which would bring in additional funding.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. However, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or the sustainability, scalability, or replicability plans. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Regents of the University of California at San Francisco

April 18, 2018

Stephanie Peterson (Mathematica Policy Research)
Shannon Heitkamp (Mathematica Policy Research)
Shivani Reddy (RTI International)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

This page has been left blank for double-sided copying.

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group or show balance between treatment and control groups for any randomized controlled trials and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

The Regents of the University of California at San Francisco, in partnership with the University of Nebraska Medical Center, used funding from HCIA R2 to create the Dementia

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The Regents of the University of California at San Francisco received an extension through February 2, 2018.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

Care Ecosystem, a program to develop, improve, and test a new model of dementia care that would address the unmet needs of patients and caregivers in the current fee-for-service (FFS) payment structure. While most dementia care today is crisis-oriented and reactive, this model emphasizes continuous and personalized care through telephone-based support and education for caregivers and patients (dyads). The primary point of contact for patients and families is a care team navigator (CTN). This program is designed to help Medicare or Medicaid beneficiaries age 45 and older with a diagnosis of dementia and their caregivers (1) improve their satisfaction with dementia care, (2) reduce caregiver burden, (3) prevent emergency-related health care costs, and (4) keep those diagnosed with dementia in the community longer.

Major changes to the Dementia Care Ecosystem program include the recruitment approach and the intensity, frequency, and duration of the services delivered. The awardee changed its recruitment approach several times. During the second year of the cooperative agreement, the University of California at San Francisco focused on outreach with additional external partners, such as a private senior care franchise and external practices, as well as community outreach to minority populations. However, this approach resulted in fewer enrolled dyads than anticipated, so the awardee further modified its approach during the third year by focusing recruitment within the University of California at San Francisco and University of Nebraska Medical Center systems, which required less administrative burden than external providers. The awardee also continued its efforts to enroll minority populations. During the third year of the cooperative agreement, the awardee also made changes to the intensity, frequency, and duration of services delivered, including (1) developing Dementia Care Ecosystem Lite and (2) graduating program participants. The Dementia Care Ecosystem Lite was rolled out for dyads that requested less frequent or intense contact from the CTNs. These dyads tended to be well-connected to resources beyond the intervention and enrolled in the program because they were interested in participating in research. Due to staffing needs, the awardee started graduating patient-caregiver dyads in the third year. Awardee leaders chose to graduate dyads with low acuity scores at baseline, although they realized this was not an ideal method.[4] After dyads graduated, they no longer received services from the CTNs.

Key program characteristics are shown in Table I.1.

[4] Acuity was measured at baseline using the Zarit burden score.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	To develop, improve, and test a new model of dementia care that would address the unmet needs of patients and caregivers in the current FFS payment structure.
Major innovation	Providing telephone-based supportive care and education for caregivers and patients, as well as medication consultation and support in planning future medical, financial, and legal decisions.
Program components	**Patient navigator or health coach.** Via telephone, the CTNs oversee care, link participants with resources, and triage questions about medical decision making to the appropriate professionals. **Care coordination.** The clinical team trains, supervises, and provides expert advice to the CTNs, intervening when specialized attention or guidance for medical decision making is needed. Clinical teams consist of a nurse, a pharmacist, and a social worker. **Medication management and adherence.** Through a computerized dashboard system, all participants receive a medication review by a pharmacist at enrollment. Medication reviews can also be triggered at the request of a clinician or a CTN or by an automated alert from the dashboard monitoring system. **Patient and family education.** Education and support resources are targeted to the patient's needs and stage of dementia. Resources include legal, financial, and medical planning, as well as behavior management and safety planning for caregivers. **Health information technology.** The dashboard clinical workflow management tool consists of (1) a caregiver module, (2) a medication module, (3) a decision making module, and (4) functional monitoring (in design phase). The dashboard also includes scheduling and data collection tools. A caregiver portal was also piloted in Year 3.
Target population	Medicare or Medicaid beneficiaries age 45 and older with a diagnosis of dementia and their caregivers, including underserved populations
Theory of change/ theory of action	The University of California at San Francisco (UCSF) hypothesizes that giving patients and caregivers personalized preventive care that is provided over the phone and supported by innovative technology should reduce the incidence of medical emergencies, prevent unnecessary ED and hospital use, and delay nursing home placement. UCSF believes these outcomes should result in overall cost savings to health care systems and improved quality of life for patients and families.
Payment model	New FFS payment, value-based payments, bundled or episode payment, capitated payment for care management and coordination services
Award amount	$9,990,848
Effective launch date[a]	3/31/2015
Program setting	By telephone (calls usually take place in the participants' home but sometimes in other settings, such as during a caregiver's lunch break from work)
Market area	Rural, urban, and suburban
Market location	CA, IA, NE
Target outcomes	• Improved caregiver perception of patient's quality of life • Heightened caregiver satisfaction with module services • Reduced caregiver burden • Reduced caregiver depression • Decreased (1) ED visit rate and costs, (2) hospitalization costs, (3) ambulance utilization and costs, (4) nursing facility costs, (5) prescription drug costs, (6) use of high-risk medications and other potentially inappropriate medications, and (7) percentage of patients with an adverse drug event

[a]After the initial planning period, the awardee's program became operational as of this date.

CTN = care team navigator; ED = emergency department; FFS = fee-for-service.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, the University of California at San Francisco enrolled 112 percent (or 512 participants) of its revised enrollment goal by the end of the three-year cooperative agreement. Although the awardee had recruitment and enrollment challenges throughout and had to lower its target several times from the original target of 1,400 treatment participants, it received a no-cost extension and seemed to have enrolled enough people early enough in the program so that their exposure to the intervention would make it likely that we would see impacts by the end of the program. Second, the awardee delivered intervention services consistent with its planned approach of implementing an agile program, or one that was iterative and continually updated based on input from staff members as well as caregivers and providers. Although the planned mode and quality (that is, the core component of the CTNs providing support and education through the telephone) remained largely the same throughout, the frequency, intensity, and duration of the delivered services varied based on caregiver and patient needs and preferences as well as resource availability. Third, the awardee's ability to recruit and hire qualified staff was for the most part successful, as was retaining most of them throughout the cooperative agreement. The awardee also provided effective, comprehensive training to staff in its service delivery protocol as well as other professional development and wellness (for example, stress reduction) trainings. However, the awardee struggled to engage providers in the program. Fourth, once participants were successfully recruited and enrolled in the program, the awardee engaged them in a timely manner and retained most of them for the full period of enrollment. Finally, implementation staff reported that the program had a positive effect on the delivery of care. Leaders reported that the intervention seemed particularly successful for caregivers, which they attributed to the caregiver's relationship with the CTN.

Impact evaluation. In our interim analysis, impact estimates provide no evidence that the dementia care program lowered Medicare expenditures, hospitalizations, or ED visits. Program effects would need to be considerably larger than 20 percent to be confident of detecting a difference in this sample. The conclusions from the impact findings should be regarded as preliminary because the treatment group consists only of beneficiaries who entered the University of California at San Francisco's program prior to November 2016. Our next analysis will include all 539 participants in the University of California at San Francisco program, which will improve our precision and statistical power somewhat. However, it will still be the case that unless estimated program impacts are large the impact estimates will not be statistically significant.

Payment model. The awardee will continue providing program services by relying on FFS reimbursement using Medicare's new care management and care planning codes, introduced in January 2017. During the third year of the cooperative agreement, the awardee continued to develop a value-based model. In addition, the awardee hired a research fellow to analyze the cost of the Dementia Care Ecosystem to support its payment model development. As part of the value-based payment model development, the awardee secured an agreement with an accountable care organization (ACO) to provide services to eligible Medicare beneficiaries. The

Next Generation ACO model, which optionally allows for capitation, also offers experienced ACOs financial arrangements with levels of risk and reward.

Sustainability plans. The University of California at San Francisco received a five-year research grant from the National Institute on Aging to partially sustain the program at the awardee's site after the end of the cooperative agreement. The awardee reported plans to use the five-year award to collect longitudinal program data necessary to show the benefits of the program and engage payers. The University of Nebraska Medical Center had not secured funding to sustain the program, but was working with two potential partners to fund an expanded version of the program for any patient with significant comorbidities. The awardee also reported progress in expanding the program to additional sites. It was working with consultants to document and standardize the program to ensure fidelity of the core components at implementing sites.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source was a survey of non-clinician staff on their perceptions of the program's effects on care delivery. The non-clinician staff survey, which was fielded from July 2016 to October 2016, achieved a response rate of 96 percent and had a sample size of 26. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we concluded that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

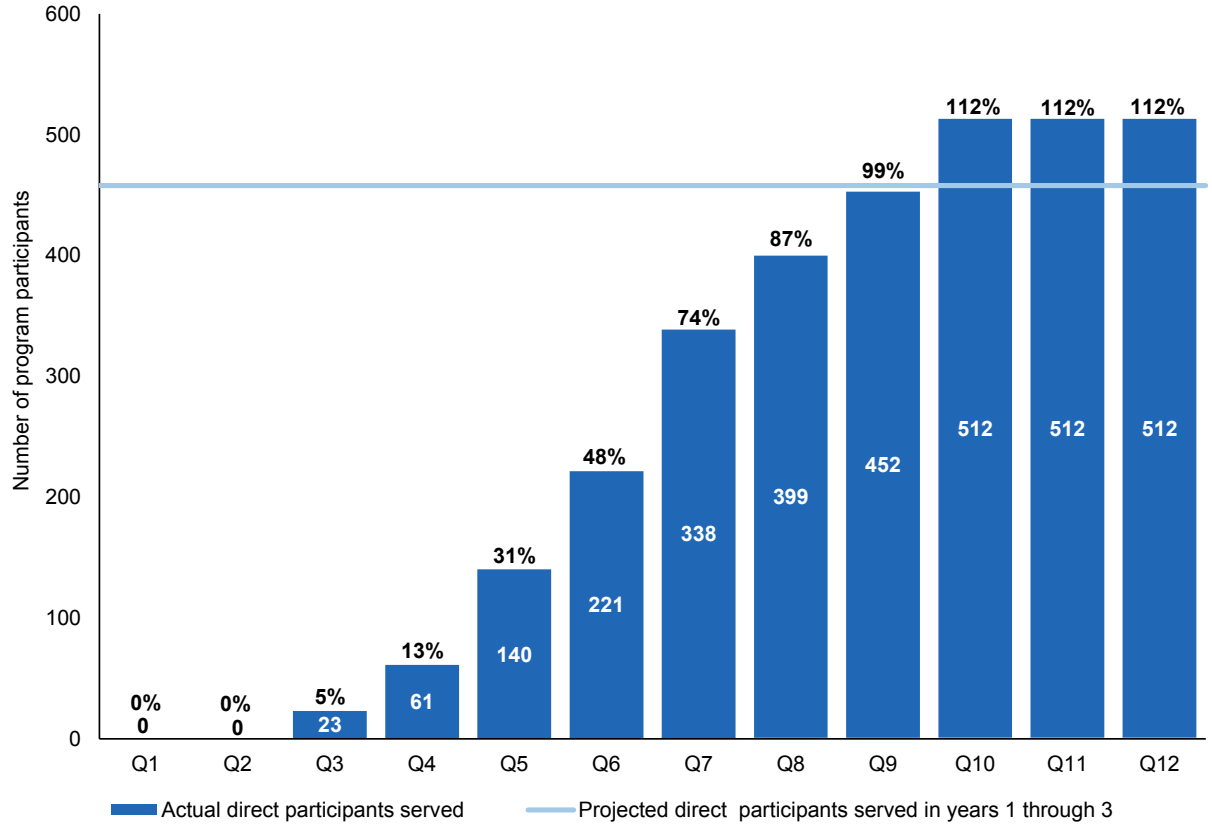
The target population consisted of Medicare or Medicaid beneficiaries age 45 and older with dementia at any stage (mild, moderate, or advanced), regardless of dementia type. Program staff also identified a primary family caregiver, who was required to be involved in the intervention. Permanent residents of nursing homes were excluded from participating. Participants were recruited from a variety of sources. The majority of participants were recruited through either self-referral or the health systems of the University of California at San Francisco and the University of Nebraska Medical Center. Following enrollment, dyads were randomized to a treatment or control group. Members of the control group received care as usual and were connected to community resources by a research coordinator (RC), who spoke with them over the phone.

Program leaders originally planned to recruit 2,100 participant dyads (1,400 for the treatment group and 700 for the control group) by the end of the three-year cooperative agreement. However, due to slower-than-expected recruitment, they lowered their projections to 844 participant dyads (458 for the treatment group and 386 for the control group). As with previous years, the University of California at San Francisco and the University of Nebraska Medical Center modified their enrollment approach in the third year by adding additional referral sites. The University of Nebraska Medical Center originally planned to obtain a large portion of referrals from area agencies on aging but these organizations did not end up providing the referrals as expected. In the final year, the University of Nebraska Medical Center opened up the program to referrals from its geriatric clinic and from the UCSF Fresno Alzheimer and Memory Center. The awardee also continued to increase its outreach to minority communities.

b. Evidence of enrollment effectiveness

The awardee reported that it enrolled 512 direct participants from March 2015 (when it launched its program) through February 2017, which represents 112 percent of its final three-year projection (Figure II.1). The awardee reduced its projected number of participants in the treatment group from 1,400 to 458 in Year 3. When measured against the three-year projection set at the end of Year 1 (enough time for the awardee to gain a better understanding of the number of eligible particpants but before the awardee began planning for the end of their cooperative agreement), the awardee met 73 percent of its projection. The University of California at San Francisco started enrolling caregivers and patients with dementia in the third quarter of the first year of the cooperative agreement, enrolling 41 dyads by the end of Year 1. In the second year of the cooperative agreement, recruitment continued to be more difficult than anticipated; however, changes in recruitment strategies proved to be successful and the awardee was able to enroll 405 dyads by the end of Year 2. During the third year of the cooperative agreement, program staff continued their efforts to engage referral providers and took advantage of connections in the community and within their own health systems to boost enrollment. Although the majority of participants were likely enrolled long enough to see some impact from the intervention, the awardee noted that the needs of dementia patients and their caregivers change as the disease progresses, so the frequency, intensity, and duration of services will also change. Therefore, the ideal exposure to the intervention would be ongoing.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee reduced its projected number of participants in the treatment group from 1,400 to 458.

c. Barriers and facilitators associated with enrollment effectiveness

The University of California at San Francisco's progress in meeting its three-year enrollment goal was influenced by several factors. This included challenges in recruitment and enrollment, such as getting external organizations and providers to help with referrals. The slower-than-expected enrollment created additional burdens on staff, which led the awardee to hire additional staff to help with recruitment and enrollment, including bilingual staff to recruit underserved populations. In addition, the rollout of the online patient-caregiver portal and the functional monitoring components of the program—though not core components of the program—was unsuccessful due to limited staff resources, technical difficulties, and a lack of interest from the dyads. However, the staff continually referred to the program as "agile," meaning that there was flexibility in the program management and design that helped the team overcome challenges and ultimately achieve success in meeting program goals.

Throughout the cooperative agreement, the awardee struggled to recruit and enroll participants in the program. Some initial recruitment strategies that program leaders thought

might yield many referrals produced only modest results (for example, relying on area agencies on aging to provide a large number of referrals). Staff therefore began to use other strategies, including maximizing media coverage and outreach at local health and wellness events, which proved successful. Program staff were surprised at how time-consuming recruitment efforts were, so program leaders hired additional staff to help mediate limited recruiting resources.

During Year 3, the program added additional referral sources to help with recruitment challenges. This included two entities that were internal to the awardee's health systems: (1) a geriatric clinic at the University of Nebraska Medical Center and (2) the UCSF Fresno Alzheimer and Memory Center. Partnering with these sites allowed program staff to engage and educate providers more successfully because there was already a level of familiarity with the staff and the program. The providers would tell potential participants about the study when they came in for their visits. Because they already knew and trusted their provider, program staff said that the dyads were more likely to join the study. Partnering with internal sites also allowed staff to meet with potential participants before or after doctor's appointments and to conduct face-to-face recruitment, which the RCs believed led to increased enrollment. Finally, partnering with the UCSF Fresno Alzheimer and Memory Center allowed additional access to the Latino community. The staff perceived this internal site as less difficult to work with than some community agencies in which there was more of an administrative burden. Nevertheless, the majority of recruitment came from the internal medical systems, including the UCSF Memory and Aging Center, or self-referrals. This resulted in a disproportionately white, high-income, and English-speaking sample, rather than the low-income, ethnically diverse, underserved population that the awardee had intended to reach. The staff believed that an underserved population would have fewer available resources to access and therefore would benefit more from the program.

> "I'll try to meet [eligible participants] in person for fifteen or twenty minutes while they're sitting in the waiting room so that they meet me and know what I look like. And a lot of times it's even just chatting about how was your drive and how's the weather. But I think it helps them feel like you're a real person and that changes my relationship with families in phone calls after that."
>
> —Care team navigator

2. Delivery of program services

a. Description of and changes to service delivery model

The Dementia Care Ecosystem program provides care management and caregiver support by telephone through the CTNs, who also link participants with any needed community resources. The care management provided by the CTNs, who are unlicensed, includes telephonic support for dementia patients and their caregivers and triaging of more advanced needs to the Dementia Care Ecosystem clinical team. Caregivers also have access to educational resources for legal, financial, and medical planning, as well as dementia care delivery from the CTN. The multidisciplinary Dementia Care Ecosystem clinical team—consisting of a nurse, a pharmacist, and a social worker—intervenes when participants need specialized attention or guidance for medical decision making. In addition, all participants receive a medication review by the team pharmacist at enrollment. Medication review can also be triggered at the request of a clinician or a CTN or by an automated alert from the computerized dashboard system used by program staff. The Dementia Care Ecosystem program aimed to improved caregiver perception of a patient's quality of life; reduce caregiver burden and depression; and reduce costs and utilization of the

ED, ambulances, nursing facilities, and prescription drugs. Continued support from the CTN was the essential component for achieving the program's desired outcomes.

Although the overall service delivery model was standardized in its core program elements—including the composition of the clinical team and the emphasis on the CTN being the point of contact to provide caregiver support and education throughout the implementation—the program team refined the model over time based on user inputs. In Year 3, this refinement involved developing the Dementia Care Ecosystem Lite program, which modified the frequency and intensity in which services were delivered for some dyads in order to continue to engage them. The Ecosystem Lite program was designed for patients and caregivers who requested fewer check-in calls, were not responsive to the CTNs phone calls, or had low acuity scores. Often times, the CTNs no longer called this group of participants at regular intervals, unless requested by the dyad. Sometimes, they communicated by mailing relevant resources as needed.

As the University of California at San Francisco was approaching the end of the cooperative agreement, it also modified the duration of services provided. Originally, the awardee had planned to include participants "indefinitely" in the program, with the understanding that support for dementia patients and caregivers would change over time due to the progression of the disease. However, during the third program year, staff began to "graduate" some program participants with low acuity scores due to constraints in available resources. Patients were considered graduated when the CTN provided them with a final call explaining graduation. Staff said that those who graduated were generally grateful to be a part of the study and sad to lose the relationship they had formed with their CTN.

b. Evidence of service delivery effectiveness

The awardee was successful in its service delivery implementation effectiveness. It delivered its core program components consistent with its planned approach and was mostly successful in hiring and retaining staff as well as providing staff with a comprehensive training program that involved problem-solving, sample scripts, and background education. Slower-than-expected recruitment and enrollment caused the program team to need a larger staff. There were some delays in hiring more staff. In addition, the awardee experienced difficulties with launching and maintaining the patient portal and the functional monitoring module. However, the technological components of the service delivery model were not critical to achieving the program's desired outcomes. Once participants were successfully recruited and enrolled in the study, the program staff successfully engaged them in a timely manner.

Delivery of intervention services. The awardee delivered intervention services consistent with its planned approach of implementing an agile program that was continually updated based on input from staff members as well as caregivers and providers. As mentioned above, the planned mode and quality of the individualized services provided to the dyads over the telephone remained consistent. The frequency, intensity, and duration of the delivered services varied based on caregiver and patient needs and preferences as well as resource availability. In addition, two aspects of the program were not rolled out successfully: (1) the patient portal and (2) the functional monitoring module. Changes to the service delivery model did not affect the awardee's ability to provide participants the core components of the service delivery model or the intended period of exposure to the intervention.

Staffing and training. The awardee's ability to recruit and hire qualified staff was mostly successful. However, in the second and third year of the cooperative agreement, there were some delays in hiring additional staff to address the recruitment and enrollment challenges. According to metrics from the implementation and monitoring contractor, the awardee was below its projections for staff hired for the project by 19 percent and 9 percent in the 11th and 12th quarters, respectively. For the most part, retaining staff throughout the cooperative agreement was successful. However, there was some staff turnover. Because the core implementation staff was young, it was common for the CTNs to leave the program and return to school full-time. In Year 3, the University of California at San Francisco focused on retaining CTN and RC staff who were concerned about job security after the cooperative agreement ended.

> "We allow [junior staff] to participate in lectures or shadowing [of senior staff and providers] so that they feel like their time with our team is valuable to them on a personal level, both in their work and just work environment."
>
> —Program leader

Engagement of program participants. Once participants were successfully recruited and enrolled in the program, the awardee engaged them in a timely manner, in a meaningful way, and mostly for the full period of the intervention. Staff reported that participants who graduated from the program during the third year often stated how they would miss the interactions with their CTN. The staff said they were able to provide participants with individualized services as needed and requested. They said they also worked hard to maintain the survey response rate of the control group by offering $25 gift cards.

> "I think being able to problem-solve for them gets [the dyads] engaged. They call me back if anything comes up."
>
> —Care team navigator

Engagement of clinical providers. Engagement of clinical providers centered on communications between the CTNs and the primary care providers about patient care. Throughout the cooperative agreement, the awardee struggled to engage primary care providers outside the networks of the University of California at San Francisco and the University of Nebraska Medical Center.

c. Barriers and facilitators associated with service delivery effectiveness

The awardee faced some barriers to delivering program services in the third year, including limited resources due to staff turnover and continued troubles with engaging providers. In addition, the University of California at San Francisco was unable to implement two minor aspects of the Dementia Care Ecosystem: (1) the online portal and (2) the functional monitoring module. However, neither component was critical to the success of the program. Facilitators in the third year included training and professional development courses to engage staff and "warm handoffs" to engage program participants and to address issues with staff turnover.

Overall, staff was hired and retained in a timely manner. However, during the third year, the CTNs and RCs faced challenges due to staff turnover, which was the result of the uncertainties in job security and future program funding. The staff turnover increased the burden on remaining staff members because they were required to take on additional job responsibilities. In our interviews, some CTNs and RCs reported decreased job satisfaction. Leaders worked to retain less-experienced CTNs through the end of the cooperative agreement by providing professional

development, including helping them to define their career goals. The change in enrollment strategy continued to take more time than planned and the CTNs and RCs perceived their workload to be very heavy during the last six months of the program due to an increased push for enrollment in order to reach the target goal. A few RCs believed that they could have enrolled more participants if they had more time.

In addition, staff turnover was a strain on resources, especially because the dyads had to be transitioned to a new CTN. Replacing the CTN was difficult and transitioning the former CTN's dyads to a new CTN was challenging for two reasons: (1) it took time and effort to transition the dyads while the current CTNs were already stretched for time and (2) the CTNs had to build trust with their new dyads. Because 79 percent of the staff reported increased feelings of burnout, awardee leaders started to hold stress management trainings.

The University of California at San Francisco worked to engage program participants in the third year despite the staffing barriers. For example, to ease the transition of dyads between CTNs due to staff turnover, the awardee standardized using a warm handoff: the leaving CTNs would have a phone contact with their dyads, explain when they were leaving, discuss a comprehensive care plan that they had developed for the dyad, and then introduce the CTN who was assuming care. This process helped make the transition more personal.

Providers outside the hospital systems were not engaged with the program despite efforts to educate them. In addition, for privacy reasons, many external providers did not want to provide patient contact information to the program staff who were responsible for contacting potential participants about the project. Even when outside providers had patients who were enrolled in the program, the staff found it difficult to engage them due to their busy schedules and relative disinterest in the program. The staff attempted to contact providers in many ways—such as by fax, email, and phone. Program leaders also directly contacted some providers. In addition, the CTNs worked with caregivers to empower them to engage with the dementia patient's provider directly (for example, by bringing in a list of current medications to discuss with the provider and any potential changes recommended by the program staff).

Because of the competing demands of enrollment and providing care, getting the online portal up and running was not a top priority for the awardee, according to leaders. Another barrier to the portal's limited rollout was slow uptake with the dyads. The online portal was designed to contain educational materials and a forum where staff could post information. It was also supposed to provide caregivers an opportunity to engage with other caregivers as well as provide secure communication with the CTN. According to one CTN, the dyads were generally not interested in learning about the portal because they were "not good at technology."

The functional monitoring module was never rolled out beyond a trial period in Nebraska, in part because of barriers to engaging the Dementia Care Ecosystem population with technology. The elderly population was not interested in using the module, according to the awardee. In addition, the CTNs could not accurately troubleshoot technological issues for dyads over the phone.

C. Assessment of perceived program effects on the delivery of care and outcomes

Based on interviews, staff believed the Dementia Care Ecosystem was successful at achieving its goals of decreasing caregiver burden, increasing patient quality of life, and providing continuous care. According to the staff surveys, 71 percent of respondents reported that the Dementia Care Ecosystem was effective at meeting its stated goals. Awardee leaders noted that the intervention was particularly successful in lowering caregiver burden, increasing caregiver self-efficacy, and improving caregivers' perceptions of the patient's quality of life. In addition, staff noted that the intervention seemed particularly successful for those dyads that were not already well-connected to resources. Staff attributed these improvements to the relationships with the CTNs. They stated that caregivers really appreciated the relationship with the CTN and the access to the dementia specialists. In addition, telephone-based care allowed dyads that might normally be harder to reach in person to participate.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. The awardee's recruitment and enrollment challenges could have implications on the ability to see expected outcomes. As discussed previously, the continual challenges to recruit participants caused awardee leaders to lower their target treatment enrollment number several times throughout the study (from 1,400 for treatment participants originally to a revised target of 458), and ultimately resulted in a lower sample size of dyads. Many of the originally planned recruitment strategies, such as referrals from area agencies on aging, were not successful. Because the program enrolled a sample that was disproportionately economically advantaged rather than disadvantaged, the potential impact on outcomes may be somewhat diluted. However, once dyads were enrolled in the study, the staff were able to successfully provide the core services and continually engage the dyads throughout the implementation period. In addition, the staff reported that enrolled participants said they found the services helpful.

Two modifications to the program may make it difficult to assess impacts on the core outcomes. First, participating dyads received different levels of the intervention based on the patient acuity scores or if they were enrolled later in the intervention and had less time to receive the Dementia Care Ecosystem services. As discussed above, patients with a low acuity score at baseline were graduated from the program because of resource constraints. Second, some dyads were enrolled in Dementia Care Ecosystem Lite, which involved fewer contacts from the CTN. The dyads that were transitioned to the less-intensive program originally received the full intervention, and therefore they are counted in the enrollment number. But because they are counted toward the total dyads enrolled, these enrollees may potentially mask some of the effects of the program.

Finally, the Dementia Care Ecosystem was never fully integrated with primary care. The CTNs attempted to empower caregivers to work with primary care providers, but the CTNs and primary care providers were not consistently working together to provide care for the dyads. Without primary care integration, the impact of medication reconciliation and care coordination

may be limited. For example, medication reconciliation by the Dementia Care Ecosystem clinical team could be undone by the primary care provider if the primary care provider continued to prescribe the medication. In addition, the Dementia Care Ecosystem clinical team may duplicate services received from the primary care provider, which would negatively impact care coordination.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of the University of California at San Francisco's dementia care program, baseline characteristics of the treatment and control groups, and preliminary impact findings for the four CMMI core outcomes.

Key findings

- Our estimates provide no evidence that the dementia care program lowered Medicare expenditures, hospitalizations, or ED visits.
- Program effects would need to be considerably larger than 20 percent to be confident of detecting a difference in this sample.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

The University of California at San Francisco's program ended August 30, 2017. The projected sample sizes in Table III.1 are based upon actual enrollment as of February 28, 2017, when enrollment ended.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: University of California at San Francisco

Evaluability domain	Response
Projected Medicare FFS population with 6 months of program exposure as of February 28, 2017	557[a]
Projected Medicaid population with 6 months of program exposure as of February 28, 2017	75[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,563
Likelihood of all-cause hospitalizations	2,454
MDE sample size requirement to detect 20% effect	
Total expenditures	391
Likelihood of all-cause hospitalizations	614
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Claims not needed

Table III.1 *(continued)*

Evaluability domain	Response
Likelihood of solid comparison group	No serious issues; proceeding with comparison group selection
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	RCT regression-adjusted analysis
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff survey
Implementation data that will be analyzed	The awardee has provided data on the following for treatment and comparison groups: Patient Quality of Life–Alzheimer's Dementia measure, caregiver Patient Health Questionnaire–9, caregiver Zarit Burden Index, caregiver self-efficacy

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

FFS = fee-for-service; RCT = randomized controlled trial.

The awardee conducted a randomized controlled trial, which allowed us to use the multivariate regression-adjusted analysis that is presented in this report. Medicare FFS claims data for both the treatment and the control group are available. However, the awardee encountered several challenges with program enrollment, lowering projections of the treatment group from 1,400 to 458 over the course of the cooperative agreement; the majority of enrollment occurred in Year 2 and Year 3. Among this reduced sample, 78 participants are enrolled in Medicare Advantage plans for which we do not have data. If we remove ineligible beneficiaries, we anticipate that the maximum total sample size of Medicare FFS participants will be 557, raising the concern that the sample may be too small to detect effects on cost and utilization. In addition, the number of Medicaid beneficiaries is too small to conduct an impact analysis.

B. Characteristics of Medicare and Medicaid participants at baseline

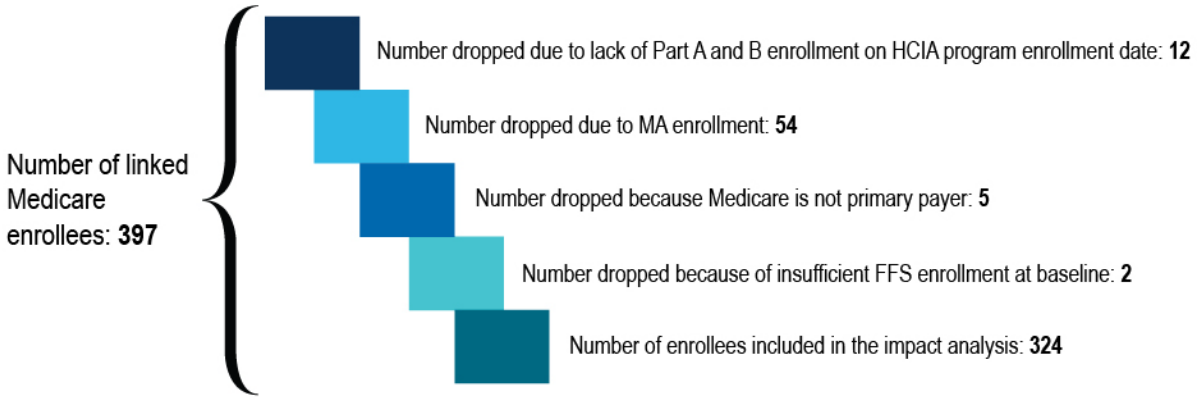
This section presents our summary of the baseline characteristics of the treatment group, which we measured during the 12 months before each beneficiary's enrollment date and entry into the Dementia Care Ecosystem program. The majority of participants have been recruited through self-referral or physician referral to the health systems of the University of California at San Francisco or the University of Nebraska Medical Center. The treatment group for this evaluation consists of Medicare beneficiaries who were randomized to the Dementia Care Ecosystem program.

The University of California at San Francisco and the University of Nebraska Medical Center began to enroll Medicare and Medicaid beneficiaries into the Dementia Care Ecosystem program in March 2015. After enrollment, participants are randomized 2:1 to either the Dementia Care Ecosystem treatment group or to the usual care control group, in which participants complete surveys only. The start date of the intervention is defined as the date of randomization. As of September 30, 2017, the awardee had enrolled 780 unique beneficiaries

into the program.[5] The Medicaid sample includes 93 beneficiaries across two sites and is not large enough to detect impact estimates on key outcomes.

In presenting the baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who had met all criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they had to have been enrolled in the awardee's program on or before November 30, 2016, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. The calendar period covered by the baseline quarters is based on the enrollment date for each enrollee and therefore varies by participant. After we excluded beneficiaries who did not meet the above criteria, a total of 486 participants were included in the analysis of baseline characteristics for this report, with 324 in the treatment group and 162 in the randomized control group.

Figure III.1 Reasons for exclusion of program enrollees from impact analyses



The Medicare FFS beneficiaries participating in the Dementia Care Ecosystem program are elderly and primarily white (Table III.2). The majority of participants in the treatment group are age 75 or older—40 percent are 75 to 84 years old, while 26 percent are age 85 or older. Only 2 percent of participants are younger than age 65. The majority of participants are female (55 percent) and white (88 percent). The original reason for Medicare eligibility was primarily age or survivor's insurance (93 percent). None of the participants were entitled to Medicare because of end-stage renal disease (ESRD). Compared with 24 percent of Medicare beneficiaries nationwide, only 7 percent of participants were originally eligible for Medicare because of a disability. Thirteen percent of participants are dually eligible for Medicare and Medicaid,

[5] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

compared with 18 percent nationally—although, in some cases, Medicaid benefits may be restricted to the payment of co-insurance and deductibles. The average hierarchical condition categories (HCC) risk score of participants is 1.34, which means that beneficiaries recruited for the Dementia Care Ecosystem program are predicted to be 34 percent more costly than the general Medicare FFS population in the first year of the program.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the University of California at San Francisco's program through November 30, 2016

	All participants (N = 324)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	8	2
65 to 74	101	31
75 to 84	130	40
85 and older	85	26
Gender		
Female	178	55
Male	146	45
Race		
White	284	88
Black	11	3
American Indian, Alaska Native, Asian/Pacific Island American, or other	16	5
Hispanic	8	2
Original reason for Medicare eligibility		
Old age and survivor's insurance	300	93
Disability insurance benefits	24	7
ESRD[a]		
Hospice[b]	3	0.93
Medicare/Medicaid dual status, percentage dual[b]	41	13
HCC score[c]		Statistic
Mean		1.34
25th percentile		0.68
Median		1.13
75th percentile		1.65

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date of randomization. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

Table III.2 *(continued)*

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.
ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3 shows baseline utilization and expenditure data for the treatment group on a common set of measures, including the four core measures from CMMI. The awardee aims to lower the total cost of care by reducing ED visits; hospital admissions; and other health services such as ambulance use, nursing home care, and potentially inappropriate high-risk medications. We examined baseline cost of care by calculating average per beneficiary per month (PBPM)[6] Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $1,041, compared with national mean Medicare spending of $864 PBPM.[7] The quarterly PBPM ranged from $929 to $1,137. The average PBPM Medicare payments for inpatient ($232) and physician ($256) services were the largest drivers of the total cost of care. Quarterly expenditures for inpatient services ranged from $200 to $298 PBPM and from $239 to $281 PBPM for physician services.

The rate of acute care hospitalizations was 294 per 1,000 Medicare FFS participants per year during the baseline year, with 23 percent of participants having at least one hospitalization during the 365 days before enrollment. The rate of ED visits was 585 per 1,000 participants per year in the baseline year, compared with the national rate of 652 per 1,000 Medicare beneficiaries annually—with the rate increasing in the third and fourth quarters compared with the first two quarters.[8]

In the baseline year, the rate of primary care visits was substantially lower than the rate of specialty services in ambulatory settings (4,952 per 1,000 Medicare FFS participants per year compared with 8,019 per 1,000 Medicare FFS participants per year, respectively). Given that the program targets a specialty population, the higher rates of specialty care may be expected. Thirteen percent of all hospital discharges were followed by a readmission in the 30-day post-discharge window in the baseline year. The percentage of hospital discharges with a 30-day readmission was lower than the national average for Medicare beneficiaries (18 percent).

[6] The months referred to in our calculations are 30-day periods rather than calendar months.

[7] See the Kaiser Family Foundation, "Medicare Spending per Enrollee, by State." Available at http://kff.org/medicare/state-indicator/per-enrollee-spending-by-residence/. Accessed October 25, 2016.

[8] See the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed May 2016.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the University of California at San Francisco's program through November 30, 2016

Types of expenditures and utilization measures	12 months before enrollment	Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	324	318	320	324	324
Average Medicare expenditures PBPM[a]					
Total	1,041 (84)	1,137 (164)	929 (118)	1,063 (134)	1,035 (137)
Acute inpatient	232 (32)	219 (55)	200 (55)	298 (79)	209 (60)
Inpatient other[b]	20 (12)	62 (44)	21 (21)	0 (0)	0 (0)
Outpatient[c]	208 (28)	236 (59)	201 (45)	166 (27)	229 (45)
Physician services	256 (19)	260 (23)	245 (25)	281 (34)	239 (20)
Home health	115 (17)	97 (24)	87 (22)	128 (28)	146 (26)
Skilled nursing facility	87 (22)	92 (38)	57 (27)	99 (46)	101 (46)
Hospice	80 (33)	134 (51)	79 (37)	49 (31)	61 (33)
Durable medical equipment	42 (18)	36 (18)	40 (21)	42 (20)	51 (25)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	294 (33)	293 (66)	238 (56)	348 (74)	296 (68)
Outpatient ED visits[e]	585 (55)	535 (83)	488 (81)	621 (116)	691 (108)
Primary care visits in any setting	5,690 (279)	5,989 (464)	5,058 (343)	5,886 (415)	5,827 (373)
Primary care visits in ambulatory settings	4,952 (236)	5,123 (350)	4,419 (292)	5,191 (349)	5,074 (287)
Specialist visits in any setting	9,323 (449)	9,608 (660)	8,851 (583)	9,177 (577)	9,654 (633)
Specialist visits in ambulatory settings	8,019 (386)	8,194 (505)	7,749 (490)	7,960 (500)	8,173 (489)
Measures of any health care utilization					
Percentage with a hospital admission[d]	23 (2)	6 (1)	6 (1)	7 (1)	6 (1)
Percentage with an outpatient ED visit[e]	36 (3)	12 (2)	11 (2)	11 (2)	14 (2)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Percentage with a 30-day readmission among all discharges	13 (4)	17 (8)	0 (0)	16 (7)	15 (7)
Percentage of participants with a readmission among all participants	3 (1)	1 (1)	0 (0)	1 (1)	1 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

C. Quality of matched comparison group

After enrollment, patient and caregiver dyads are randomized 2:1 to either the treatment or control group. Members of the control group receive care as usual and are connected to community resources by an research coordinator, who speaks with them over the phone. Because of randomization, there's no need for developing a comparison group and propensity score matching (and thus no discussion of potential comparison groups, matching variables, and so on) that we present for other awardees with quasi-experimental designs.

Figure III.2 shows the balance between the two groups in terms of standardized differences of important baseline variables, with boundaries at 0.10 and 0.25 above and below 0.[9] Standardized differences measure the difference in weighted means between the treatment group and the matched comparison group on the standard deviation scale. Overall, we observe good balance on baseline sociodemographic and health status characteristics and prior utilization and Medicare expenditures. The vast majority of baseline characteristics fall within the standardized

[9] A full balance table is presented in Appendix A (Table A.1).

difference boundary of 0.10 above and below 0. Several variables are outside of the 0.10 boundary but are inside a 0.25 boundary, a standard threshold suggested by Rubin.[10]

We note that standardized differences for four variables fall between 0.10 and 0.25 below 0 for treatment versus control beneficiaries. In all four cases, these values are lower for the treatment group than for the randomized control group:

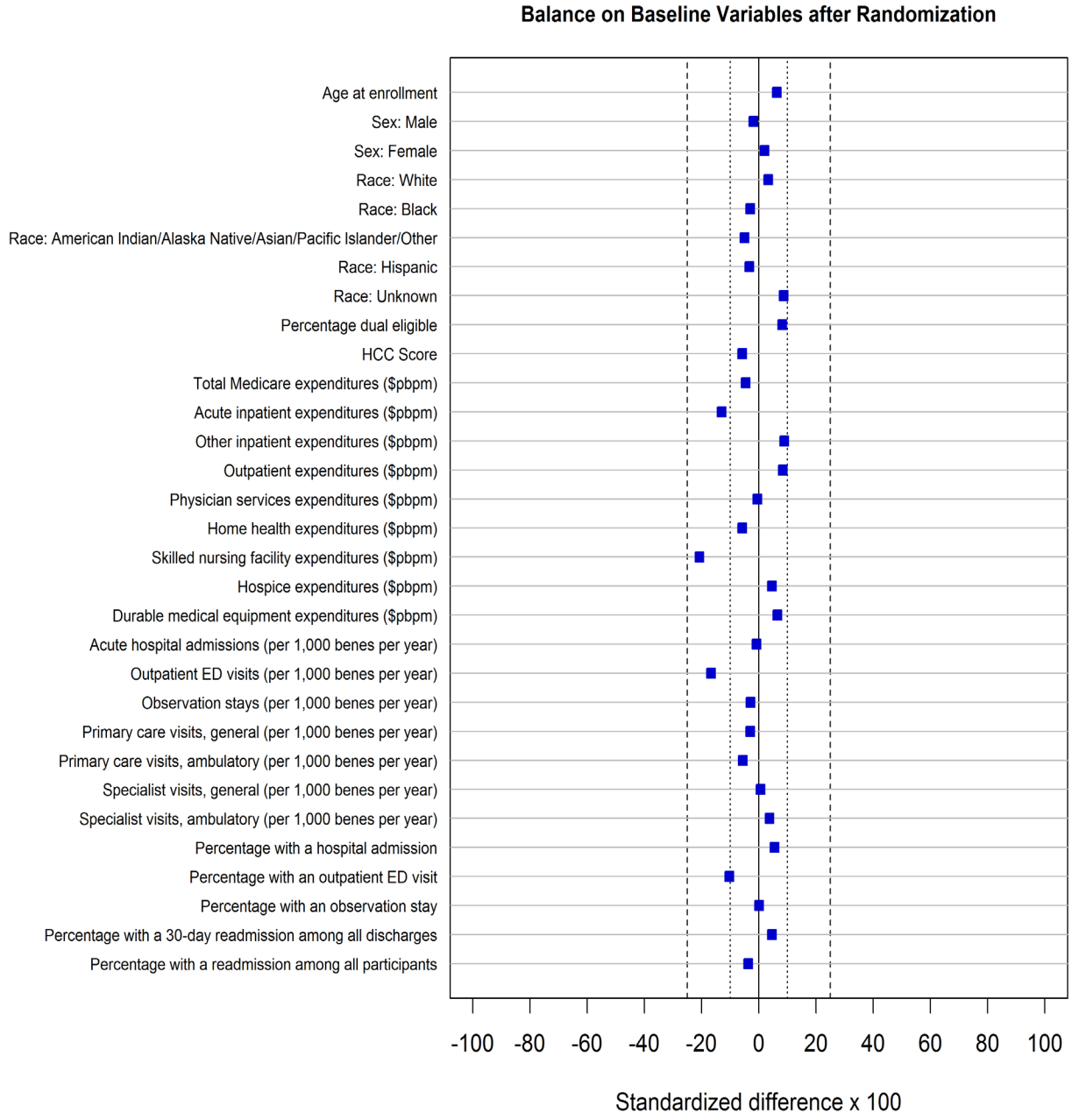
- Acute inpatient expenditures (difference = $75 PBPM)
- Skilled nursing facility expenditures (difference = $79 PBPM)
- Number of outpatient ED visits (difference = 152 per 1,000 beneficiaries)
- Percentage with any outpatient ED visit (difference = 5%)

The differences between treatment and control beneficiaries are not statistically significant, but the standardized differences are somewhat higher than our general goal of having standardized differences within a $\pm$0.10 boundary.

Given the closeness of the means for treatment and control beneficiaries on the large set of baseline characteristics, we believe that this randomized group of intervention and treatment beneficiaries will yield credible estimates of program effects. Our regression analysis will control for the remaining differences between the two groups on these baseline characteristics.

[10] Rubin, D. B. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology,* vol. 2, no. 3–4, 2001, pp. 169–188.

Figure III.2. Standardized differences of baseline variables between treatment and control beneficiaries for the University of California at San Francisco



D. Interim impact findings

The University of California at San Francisco is testing a new model of dementia care that addresses unmet needs of community-dwelling patients and caregivers in the current FFS payment structure. The program provides telephone-based supportive care and education for caregivers and patients, as well as medication consultation and support in planning future medical, financial, and legal decisions. The awardee hypothesizes that giving patients and caregivers personalized preventive care over the phone should reduce the incidence of medical emergencies, prevent unnecessary ED and hospital use, and delay nursing home placement. The awardee believes these outcomes should result in overall cost savings to health care systems and improved quality of life for patients and families.

1. Data and estimation approach

Our impact analyses includes Medicare FFS beneficiaries who were randomized into either the treatment or control arm of the awardee's program between March 31, 2015, and November 31, 2016, to allow for a minimum of six months of program exposure. We created three analytic samples with different follow-up periods: (1) 379 Medicare FFS beneficiaries who enrolled early enough to have 12 months of follow-up after randomization, (2) 486 Medicare FFS beneficiaries with one to 6 months of follow-up, and (3) 354 Medicare FFS beneficiaries with 7 to 12 months of follow-up.

Beneficiaries who died within a given follow-up period are included in the analyses for that period. We report findings for the 12-month follow-up period in this chapter; findings from the other two follow-up periods appear in Appendix B. The awardee used a 2:1 randomization scheme, which yielded 247 treatment beneficiaries and 132 control beneficiaries in the 12-month follow-up analytic sample.

Our main data sources are the Medicare Enrollment Database and Medicare FFS claims. We used Medicare enrollment data from March 31, 2014, through November 31, 2016, to identify Medicare program eligibility and beneficiary demographic characteristics. We used Medicare FFS claims from March 31, 2014, through May 30, 2017, to construct baseline and intervention period outcome measures and claims-based covariates. We allowed for three months of claims run-out and pulled the enrollment and claims data in September 2017.

Our framework is an intent-to-treat approach. We used regression models to estimate mean differences in outcomes between the treatment and control groups during the program period. The models control for all of the baseline beneficiary characteristics and baseline values of the outcomes we evaluated in Section III.C, with the exception of other inpatient expenditures, primary care visits in all settings, and specialist visits in all settings. We included baseline values of primary care and specialist visits in ambulatory settings in all models. In addition, the regression models included the following variables: original reason for Medicare eligibility, state of residence (California or Nebraska/Iowa), and maturity (months between the program start date and enrollment of each individual beneficiary).

2. Outcomes

Our first impact evaluation examines the effects of the awardee's program on a small set of outcomes (Table III.4). Because of the small sample size and imprecise estimates, we restricted our reporting to CMMI's four core measures and several related outcome measures. We examined changes between the treatment and control groups during the first 12 months of program follow-up in (1) PBPM Medicare FFS expenditures, in total and for acute inpatient care; (2) rates of hospitalization and ED visits or observation stays per 1,000 participants; and (3) likelihood of a hospital admission, ED visit or observation stay, and a 30-day unplanned readmission among all participants. In the final report, we will expand the expenditure and utilization measures that we report on and we will examine the delay in time to nursing home placement by using nursing home assessment data. We expect to have 539 participants (359 intervention and 180 control) with at least six months of program exposure for inclusion in the final report analyses.

Table III.4. Outcome measures

Outcome
Total Medicare FFS expenditures
Acute inpatient expenditures
Acute hospital admission rate
Outpatient ED visit and observation stay rate
Percentage with a hospital admission
Percentage with an outpatient ED visit or observation stay
Percentage with a readmission among all participants

ED = emergency department; FFS = fee-for-service.

3. Impact estimates

For five of the seven measures we report on, the treatment group has Medicare FFS expenditures and utilization that are 8 percent to 20 percent lower than the control group means. However, none of the differences are statistically significant at the 10 percent confidence level (Table III.5). The exceptions are outpatient ED visits and readmissions, for which we see no difference between the treatment and control groups. We observe a similar pattern of findings when we examine impacts for months 1 to 6 of follow-up and months 7 to 12 of follow-up. Given the number of participants included in this interim analysis (247 treatment and 132 control beneficiaries), we do not have a sufficient sample size to be confident of detecting a 20 percent or smaller effect of the program on total Medicare FFS expenditures or likelihood of any health care utilization (as noted in Table III.1 and reinforced by the wide 90 percent confidence intervals displayed in Table III.5). The lower end of the 90 percent interval shows that point estimates would have had to be 24 percent or greater of the predicted treatment group mean in order to be statistically significant.

Table III.5. Estimated impact of the University of California at San Francisco's program on Medicare FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a 12-month follow-up period

Expenditures and utilization measures	Randomized treatment group predicted mean (N = 247)	Randomized control group (N = 132)	Impact (treatment – control)	*p*-value	90% confidence interval
Average Medicare expenditures PBPM[a]					
Total	$1,397	$1,518	-$122	0.575	(-480, 236)
Acute inpatient	386	484	-98	0.395	(-288, 92)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[b]	373	447	-74	0.407	(-222, 73)
Outpatient ED visits[c]	746	747	-1	0.991	(-215, 212)
Measures of any health care utilization					
Percentage with a hospital admission[b]	25%	30%	-5%	0.366	(-0.12, 0.04)
Percentage with an outpatient ED visit or observation stay	37%	44%	-6%	0.267	(-0.15, 0.03)
Percentage with a readmission among all participants	4%	4%	0%	0.969	(-0.04, 0.03)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment data as of September 30, 2017.

[a]Total Medicare expenditures were calculated from all claims for each participant with at least one eligible day during the baseline year and quarter.

[b]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[c]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

4. Sensitivity analyses

We verified the robustness of our impact estimates by conducting two sets of sensitivity analyses: (1) extending the baseline period from one year to two years and (2) top-coding outcomes to account for outliers with extreme values and re-estimating the regression models. Neither of the sensitivity analyses produced results that differed substantively from those presented in this report. Due to the small number of participants included in this intervention, no subgroup analyses are planned.

5. Discussion: The University of California at San Francisco

The awardee expected to reduce the incidence of medical emergencies, prevent unnecessary ED and hospital use, and delay nursing home placement, which should result in overall cost savings. Greater than halfway through the intervention period, we do not have any evidence that the program has resulted in lower Medicare expenditures, hospitalizations, and ED visits. However, true program effects would need to be considerably larger than 20 percent for us to be

likely to find a statistically significant difference. Our next analysis will include all 539 participants in the awardee's program, which will improve our precision and statistical power somewhat. However, it will still be the case that unless estimated program impacts are large the impact estimates will not be statistically significant.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

The awardee intends to continue to provide program services by relying on FFS reimbursement using Medicare's newly introduced (in January 2017) care management and care planning codes. During the third year of the cooperative agreement, the awardee continued to develop its value-based model and hired a research fellow to analyze the cost of the Dementia Care Ecosystem to support the payment model development. As part of the value-based payment model development, the awardee secured an agreement with an ACO to provide services to eligible Medicare beneficiaries. The Next Generation ACO model, which optionally allows for capitation, also offers experienced ACOs financial arrangements with higher levels of risk and reward.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

In the future, the awardee planned to support the Dementia Care Ecosystem intervention through two payment mechanisms. First, the University of California at San Francisco had begun billing Medicare under the complex chronic care management codes that were released in January 2017, as well as the advanced care planning codes (G0505 code for care planning and dementia). Second, the awardee partnered with Allina Health, a Medicare Next Generation ACO that provides care for about one-third of the Medicare population in Minnesota and western Wisconsin. The goal of this partnership was to ensure that the program worked in risk-sharing organizations in addition to strictly FFS organizations like the awardee.

C. Status of the payment model

In the third year of the cooperative agreement, the awardee focused on implementing new chronic care management billing codes, finalizing the cost of the model, and evaluating the return on investment as claims data became available.

The awardee is using claims data to determine the cost of care delivery and the amount reimbursable by the new chronic condition codes. The University of California at San Francisco hypothesized that if the reimbursement from the new chronic condition codes could cover the cost of care delivery (or at least a substantial portion), the Dementia Care Ecosystem would be an attractive model for health care systems.

For the value-based model, the awardee is working to determine the cost of care delivery in conjunction with the amount reimbursable by the new care conditions codes. To determine actual savings, the awardee compared claims data between the control group and the treatment group to evaluate any difference in CMS health care utilization and costs. To determine the return on investment, the awardee is subtracting the cost of care delivery from the savings.

D. Factors associated with the development of the payment model

The awardee mentioned support from the implementation and monitoring contractor as helpful in providing one-on-one consultations as well as group webinars. A faculty member at the University of California at San Francisco who formerly headed the Agency for Healthcare Research and Quality also advised the awardee on implementing the new chronic care management codes.

Barriers to the development of the payment model include trying to get the University of California at San Francisco to implement the new billing codes and making sure that the compliance office approves the codes. The difficulties are consistent across all of the operating states, and mostly refer to the time and effort it takes to implement new coding schemes. In addition, the awardee noted that quantifying the long-term outcomes of the Dementia Care Ecosystem program to support the development of the value-based model was a challenge. The length of the cooperative agreement only allowed the awardee to show 6-month and 12-month outcomes.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The University of California at San Francisco received a five-year research grant from the National Institute on Aging to partially sustain the program at the awardee's site after the end of the cooperative agreement. The awardee reported plans to use the five-year award to collect longitudinal program data necessary to show the benefits of the program and engage payers. The University of Nebraska Medical Center had not secured funding to sustain the program, but was working with two potential partners that would fund an expanded version of the program to serve any patient with significant comorbidities. The awardee also reported progress in expanding the program to more sites. It was working with consultants to document and standardize the program to facilitate fidelity of the core components at implementing sites.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, the University of California at San Francisco was looking at different funding options to sustain the program. The awardee had started conversations with several health plans about potential reimbursement mechanisms and was also applying for grants to fund the program after funding from the cooperative agreement ended.

C. Implementing the SSR plan: progress and changes

Sustainability. After realizing that it needed more longitudinal program data to show the benefits of the program and engage payers, the University of California at San Francisco applied and received a five-year research grant from the National Institute on Aging. Because the award will only cover about half of the dyads, if approved, the awardee reported plans to sustain dyads with a caregiver burden level that was higher than average at baseline.

Although the University of Nebraska Medical Center had not secured funding to sustain the program, it reported that two potential partnerships would allow a variation of the program to be sustained. The two potential partners, Methodist Hospital and Blue Cross Blue Shield in Nebraska, reportedly expressed high interest in the program. If the partnerships occur, site leaders believed that the target population would likely change to include any patient with a significant comorbidity—such as coronary artery disease, congestive heart failure, emphysema, type 2 diabetes, or patients with more than one ED visit per six months. Program services may also change to focus only on patients, rather than patients and their caregivers.

Scalability. In the third program year, the University of California at San Francisco also reported progress in expanding the program to more sites. The awardee listed multiple new sites that were implementing the program in the third program year, including two sites in San Francisco funded by the Administration for Community Living, the awardee's specialty neurology clinic at the Memory and Aging Center, the Curry Senior Center, and Alina Health. The University of Nebraska Medical Center was in conversations with another site, the Veterans Affairs Medical Center in Omaha, to also implement the program. The awardee hired consultants to help document and standardize the program to facilitate fidelity of the program's core components at implementing sites, while also encouraging adaptations based on specific needs of the site. Many of the new implementing sites already had versions of the program implemented, awardee leaders said, so the program is helping those sites based on the community's needs.

Replicability. The University of California at San Francisco did not report plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

The University of California at San Francisco reported that the 5-year funding from the National Institute on Aging would facilitate long-term sustainment of the program by allowing the awardee to collect longitudinal data on program outcomes. Before the awardee had secured the grant, however, the lack of secured funding challenged its ability to retain program staff.

In addition, the awardee had specific challenges with implementing and sustaining the program at Veterans Affairs Medical Centers. One program leader explained that Veterans Affairs sites take a long time to negotiate and implement programs and that turnover in organizational leaders that occurred at one potential implementing site may delay implementation even further.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or SSR plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

The conclusions from the impact findings should be regarded as preliminary because the treatment group consists only of beneficiaries who entered the awardee's program prior to November 2016. In the next analysis, we will include all beneficiaries who enrolled prior to February 2017, to allow for six months of program exposure prior to the end of the program on August 30, 2017. With the final study sample, we will ensure balance between the randomized treatment and control groups and re-estimate the impact regression models to arrive at final impact estimates. The larger sample can be expected to result in lower standard errors and therefore provide more precise estimates of program effects. The model will be estimated for a longer period of program operations, permitting us to estimate the effects, if any, of program maturity, as well as longer exposure to the program for individual beneficiaries. We will also perform more extensive sensitivity and robustness checks on all results.

In addition, we will estimate program impacts in a Bayesian framework, permitting the estimation of the program's impact to borrow strength from estimates from other time periods and outcomes and also allowing us to estimate the probability that a program impact exceeds a specified statistical threshold. We will expand our analyses of outcomes to include delay to nursing home placement by using assessment data available for Medicare beneficiaries who enter a custodial care facility. We will also evaluate the utility of awardee-collected data on participant quality of life and caregiver depression, burden, and self-efficacy. We will provide CMS with final impact results as they become available in the upcoming year.

This page has been left blank for double-sided copying.

APPENDIX A

POST-MATCHING DIAGNOSTICS

This page has been left blank for double-sided copying.

Table A.1. Matching covariate balance for University of California at San Francisco

Measure	Treated (N = 359)	Controls (N = 180)	Difference	Percentage difference	Standardized difference	*p*-value	Equivalence *p*-value
Age at enrollment	78.677 (0.424)	78.172 (0.701)	0.505 (0.790)	0.600	0.063	0.516	0.029
Proportion male	0.451 (0.026)	0.461 (0.037)	-0.010 (0.044)	-2.200	-0.020	0.829	0.005
Proportion female	0.549 (0.026)	0.539 (0.037)	0.010 (0.044)	1.800	0.020	0.829	0.005
Proportion White	0.872 (0.018)	0.861 (0.026)	0.011 (0.031)	1.200	0.032	0.728	0.010
Proportion Black	0.033 (0.009)	0.039 (0.014)	-0.005 (0.017)	-16.300	-0.030	0.746	0.010
Proportion American Indian/Alaska Native/Asian/Pacific Islander/Other	0.050 (0.012)	0.061 (0.018)	-0.011 (0.021)	-21.900	-0.050	0.595	0.018
Proportion Hispanic	0.028 (0.009)	0.033 (0.013)	-0.005 (0.015)	-19.700	-0.033	0.724	0.010
Proportion unknown race	0.017 (0.007)	0.006 (0.006)	0.011 (0.009)	66.800	0.087	0.281	0.009
Proportion dual eligible	0.139 (0.018)	0.111 (0.023)	0.028 (0.030)	20.200	0.081	0.359	0.025
HCC score	1.323 (0.050)	1.378 (0.082)	-0.055 (0.097)	-4.200	-0.058	0.545	0.031
Total Medicare expenditures ($PBPM)	999.489 (77.461)	1067.240 (151.651)	-67.750 (173.990)	-6.800	-0.046	0.658	0.043
Acute inpatient expenditures ($PBPM)	226.135 (30.191)	300.926 (83.057)	-74.791 (88.908)	-33.100	-0.131	0.303	0.222
Other inpatient expenditures ($PBPM)	22.388 (11.585)	2.939 (2.939)	19.449 (11.813)	86.900	0.089	0.239	0.001
Outpatient expenditures ($PBPM)	196.524 (25.200)	156.630 (23.358)	39.894 (34.371)	20.300	0.084	0.309	0.011
Physician services expenditures ($PBPM)	253.470 (17.558)	255.176 (31.983)	-1.706 (37.609)	-0.700	-0.005	0.959	0.016
Home health expenditures ($PBPM)	110.276 (15.695)	127.693 (23.903)	-17.417 (29.623)	-15.800	-0.059	0.532	0.028
Skilled nursing facility expenditures ($PBPM)	80.204 (20.013)	159.078 (47.102)	-78.873 (52.439)	-98.300	-0.208	0.072	0.381
Hospice expenditures ($PBPM)	71.609 (29.717)	46.112 (27.178)	25.497 (40.002)	35.600	0.045	0.581	0.002
Durable medical equipment expenditures ($PBPM)	38.883 (16.456)	18.687 (4.608)	20.197 (17.207)	51.900	0.065	0.390	0.000

Table A.1 *(continued)*

Measure	Treated (N = 359)	Controls (N = 180)	Difference	Percentage difference	Standardized difference	*p*-value	Equivalence *p*-value
Acute hospital admissions (per 1,000 beneficiaries per year)	285.197 (30.872)	290.326 (52.086)	-5.129 (61.313)	-1.800	-0.009	0.928	0.011
Outpatient ED visits (per 1,000 beneficiaries per year)	513.609 (47.746)	665.421 (77.707)	-151.812 (91.796)	-29.600	-0.168	0.082	0.209
Observation stays (per 1,000 beneficiaries per year)	75.209 (14.486)	83.333 (23.471)	-8.124 (28.306)	-10.800	-0.030	0.758	0.017
Primary care visits, general (per 1,000 beneficiaries per year)	5688.183 (267.660)	5842.588 (383.958)	-154.406 (467.069)	-2.700	-0.030	0.740	0.009
Primary care visits, ambulatory (per 1,000 beneficiaries per year)	4948.947 (227.789)	5192.588 (318.483)	-243.641 (388.360)	-4.900	-0.056	0.535	0.016
Specialist visits, general (per 1,000 beneficiaries per year)	9403.942 (429.685)	9360.518 (663.158)	43.425 (784.744)	0.500	0.005	0.955	0.006
Specialist visits, ambulatory (per 1,000 beneficiaries per year)	8166.937 (372.601)	7906.016 (555.298)	260.921 (662.797)	3.200	0.037	0.691	0.012
Proportion with a hospital admission	0.223 (0.022)	0.200 (0.030)	0.023 (0.037)	10.300	0.055	0.543	0.014
Proportion with an ED visit	0.334 (0.025)	0.383 (0.036)	-0.049 (0.043)	-14.700	-0.104	0.261	0.054
Proportion with an observation stay	0.072 (0.014)	0.072 (0.019)	0.000 (0.024)	0.300	0.001	0.993	0.004
Proportion with a 30-day readmission among all discharges	0.131 (0.023)	0.111 (0.028)	0.020 (0.036)	15.100	0.046	0.601	0.008
Proportion with a readmission among all participants	0.022 (0.008)	0.028 (0.012)	-0.005 (0.015)	-24.700	-0.037	0.695	0.018

	Chi-squared statistic	Degrees of freedom	*p*-value
Omnibus test for balance on all variables	26.073	28.000	0.569

Note: Standard errors in parentheses. Standardized difference calculated as the difference in means divided by the treatment group standard deviation. P-values come from a weighted two-sample t-test. The equivalence test comprises two one-sided tests of whether the difference between the treatment and comparison group means exceeds 0.25 standard deviations (in either direction) of that variable. If we can reject that hypothesis, we are confident that the true difference between the two populations being compared is less than 0.25 standard deviations.

ED = emergency department; FFS = fee-for-service; HCC=hierarchical condition category; PBPM = per beneficiary per month.

APPENDIX B

IMPACT ESTIMATE TABLES

This page has been left blank for double-sided copying.

Table B.1. Estimated impact of the University of California at San Francisco's program on Medicare FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a 12-month follow-up period

Expenditures and utilization measures	Randomized treatment group predicted mean (N = 247)	Randomized control group (N = 132)	Impact (treatment – control)	*p*-value	90% confidence interval
Average Medicare expenditures PBPM ($)[a]					
Total	1,397	1,518	-122	0.575	(-480, 236)
Acute inpatient	386	484	-98	0.395	(-288, 92)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[b]	373	447	-74	0.407	(-222, 73)
Outpatient ED visits[c]	746	747	-1	0.991	(-215, 212)
Measures of any health care utilization (%)					
Percentage with a hospital admission[b]	25.3	29.7	-4.4	0.366	(-12.3, 3.5)
Percentage with an outpatient ED visit or observation stay	41.1	46.4	-5.3	0.328	(-14.2, 3.6)
Percentage with a readmission among all participants	3.5	3.5	-0.1	0.969	(-3.6, 3.4)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment data as of September 30, 2017.

[a]Total Medicare expenditures were calculated from all claims for each participant with at least one eligible day during the baseline year and quarter.

[b]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[c]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table B.2. Estimated impact of the University of California at San Francisco's program on Medicare FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a one-month to six-month follow-up period

Expenditures and utilization measures	Randomized treatment group predicted mean (N = 324)	Randomized control group (N = 162)	Impact (treatment – control)	*p*-value	90% confidence interval
Average Medicare expenditures PBPM ($)[a]					
Total	1,267	1,345	-78	0.740	(-468, 311)
Acute inpatient	308	404	-96	0.437	(-298, 107)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[b]	315	372	-56	0.570	(-220, 107)
Outpatient ED visits[c]	669	768	-99	0.487	(-335, 136)
Measures of any health care utilization (%)					
Percentage with a hospital admission[b]	12.5	13.9	-1.3	0.693	(-6.7, 4.1)
Percentage with an outpatient ED visit or observation stay	25.9	29.2	-3.3	0.444	(-10.4, 3.8)
Percentage with a readmission among all participants	1.3	1.8	-0.6	0.648	(-2.7, 1.5)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment data as of September 30, 2017.

[a]Total Medicare expenditures were calculated from all claims for each participant with at least one eligible day during the baseline year and quarter.

[b]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[c]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table B.3. Estimated impact of the University of California at San Francisco's program on Medicare FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a 7-month to 12-month follow-up period

Expenditures and utilization measures	Randomized treatment group predicted mean (N = 230)	Randomized control group (N = 124)	Impact (treatment – control)	*p*-value	90% confidence interval
Average Medicare expenditures PBPM ($)[a]					
Total	1,416	1,608	-192	0.474	(-634, 250)
Acute inpatient	456	497	-41	0.780	(-283, 201)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[b]	426	487	-61	0.606	(-257, 134)
Outpatient ED visits[c]	788	776	12	0.947	(-270, 293)
Measures of any health care utilization (%)					
Percentage with a hospital admission[b]	16.2	20.1	-4.0	0.362	(-11.1, 3.1)
Percentage with an outpatient ED visit or observation stay	25.6	28.8	-3.2	0.524	(-11.6, 5.2)
Percentage with a readmission among all participants	2.3	1.2	1.1	0.457	(-1.2, 3.4)

Source: Mathematica analysis of information from awardee's finder file through November 2016 and Medicare claims and enrollment data as of September 30, 2017.

[a]Total Medicare expenditures were calculated from all claims for each participant with at least one eligible day during the baseline year and quarter.

[b]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[c]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: University Hospitals Cleveland Medical Center

April 18, 2018

Allison Steiner (Mathematica Policy Research)
Lisa Lines (RTI International)
Lee-Lee Ellis (Mathematica Policy Research)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Cleveland Medical Center received a six-month extension and will continue to provide formal program services through February 28, 2018. Cleveland Medical Center will maintain components of the program beyond February 28, 2018 (see Chapter VII).

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

A disproportionate share of spending on oncology-related care takes place in the last several weeks of life, even though evidence does not support the efficacy of these end-of-life interventions. University Hospitals Cleveland Medical Center is using funding from HCIA R2 to create Learning Individual Needs and Coordinating Care (LINCC), an intervention that provides care coordination and early palliative care in an effort to enhance the quality and experience of care while reducing the cost of care. This awardee works to achieve these goals by providing care that adheres to evidence-based guidelines, including early palliative care and proactive patient management with the support of a care coordinator.

The program was implemented at four sites in the Cleveland, Ohio, area: Seidman Cancer Center (main campus in downtown Cleveland) and three of its community satellite clinics. The awardee initially planned to implement the program at two more satellite sites, but did not do so. The program's target population includes adult Medicare and Medicaid beneficiaries who are receiving care at Seidman Cancer Center (and its community satellite locations) for complex cancers, which are late stage solid tumors or cancers with disease progression, or regionalized malignancies with complicating comorbidities, or cancers complicated by other risk factors for poor outcomes and higher spending (for example, poor social support or low socioeconomic status). Table I.1 has details. Cleveland Medical Center intended to serve 1,776 participants over the course of the cooperative agreement.

The LINCC program consisted of a clinical intervention (the primary program component) that provided care management and palliative care. LINCC nurse care coordinators worked with disease teams (doctors and nurses who specialize in treating a specific cancer) to manage patient care. The nurse care coordinators followed patients throughout the course of their treatment, linked patients to health system and community resources, and engaged and educated patients and their families. They also worked with patients to develop advance directives and promote adherence to the patient-centered plans of care, which include patients' goals, future appointments, and current medications. Two palliative care providers delivered early and ongoing palliative care to improve management of pain and other symptoms and address other domains of palliative care. The program also had two secondary interventions: a spiritual component and a pharmacy component (Table I.1)

To help identify and assess participants' needs, the LINCC program staff asked participants to complete a routine biopsychosocial patient assessment, which was administered on an iPad. The assessment included questions from several validated tools that are designed to evaluate patients' physical needs and symptoms, emotional state, and social well-being. The nurse care coordinators and palliative care providers used a patient's responses to focus their services on the patient's needs. They also encouraged the oncologists at Seidman Cancer Center to review patients' responses. The duration and intensity of the intervention varied depending on the patient's acuity and the patient's needs, as determined by the iPad assessment. Table I.1 summarizes the characteristics of the LINCC program and program goals.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The program focused on individuals with complex and advanced cancer and was intended to lower their health care costs, increase their satisfaction with care, and improve or maintain their quality of care through patient-centered care coordination.
Major innovation	Higher quality and better experience of care through adherence to evidence-based guidelines (including early palliative care and care coordination).
Program components	• Care management: A nurse care coordinator helped the participant establish a plan of care, served as the participant's point of contact and advocate, facilitated patient and family engagement and education, linked the participant and family to resources, and ensured that outpatient care was well coordinated. • Direct care provision: Participants received early and ongoing access to expert-level palliative care. • Health IT: Participants received routine biopsychosocial assessment administered on an iPad. • Spiritual intervention: A spiritual care coordinator gave spiritual and emotional support to participants with spiritual needs. • Pharmacy intervention: Pharmacy-related support offered an additional route to cost savings by developing protocols that encourage cost-effective drug use.
Target population	Adults receiving complex cancer care at Seidman Cancer Center; eligible patients include complex cancer patients; that is, patients with late stage (3 and 4) solid tumors or disease progression, patients with regionalized malignancies with complicating comorbidities, and patients with other risk factors for poor outcomes and higher spending levels.
Payment model	Cleveland Medical Center's payment model consists of PBPM payments of $160 for program services in addition to traditional FFS payments. Clinical services that are otherwise reimbursable by payers are not covered by the PBPM payment.
Theory of change/ theory of action	Nurse care coordinators identify adults receiving care for complex cancers and assess their physical, emotional, and spiritual needs. These needs are communicated to other members of the disease team. Nurse care coordinators work with patients to develop a plan of care and connect them with resources. Nurse care coordinators provide an extra layer of support that helps patients adhere to their plan of care, improve self-efficacy, and manage their physical, emotional, and spiritual needs. Better management of patient needs results in better quality of care, participant satisfaction, and appropriate service utilization, which in turn results in improved health outcomes across the continuum of complex cancer care and lower health care costs.
Award amount	$4,675,383
Effective launch date[a]	2/19/2015
Program setting	Seidman Cancer Center is a freestanding, comprehensive cancer hospital on the Cleveland Medical Center campus that houses a multitude of specialists and services for cancer care. The community satellite clinics are extensions of Seidman Cancer Center in suburban Cleveland.
Market area	Urban, suburban
Market location	Ohio

Table I.1 *(continued)*

Program characteristic	Description
Target outcomes	• Maintain or improve quality of care compared to that measured in 2013 baseline data (when available) and comparable peer group • Improve the participant-reported experience of care for a cohort of complex cancer patients by 5% from 2013 over experience of a comparable peer group • Improve the efficiency of health care delivery by reducing total cost of care for a cohort of complex cancer patients by 8% from 2013 costs • Demonstrate feasibility and sustainability of an innovative, asymmetrical, shared savings payment model to support enhanced service delivery • Decrease avoidable ED visits, hospitalizations, and 30-day hospital readmissions (These are not among the awardee's primary program goals, but these measures are required by the HCIA R2 cooperative agreement and will be examined as part of this evaluation.)

[a]After the initial planning period, the awardee's program became operational as of this date.
ED = emergency department; FFS = fee-for-service; IT = information technology; LINCC = Learning Individual Needs and Coordinating Care; PBPM = per beneficiary per month.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the three-year cooperative agreement. We based this conclusion on six factors. First, the awardee enrolled 1,331 participants—75 percent of its enrollment target—by the end of the initial cooperative agreement. Second, the awardee successfully implemented the primary intervention on time, including the patient assessments, nurse care coordination, and palliative care. There were delays, however, in implementing the program at the community satellite sites and in implementing the secondary components of the program— the pharmacy and spiritual care intervention. Third, the awardee fell short of its staffing targets, and the awardee believes that although this affected enrollment, it did not impact the quality or types of services provided to participants. Fourth, even though it was not easy to integrate the new program into the existing clinical workflow at first, the awardee made considerable progress integrating the program into Seidman Cancer Center and engaging providers. Fifth, the program successfully engaged patients by providing an extra layer of support and adjusting the intervention to their needs. Finally, participating clinicians and other implementation staff reported that the program had a positive effect on the delivery of care.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Cleveland Medical Center's LINCC program, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. Cleveland Medical Center was proceeding with participation in the CMS-sponsored Oncology Care Model (OCM), a program launched by CMS in July 2016 that involves monthly enhanced payments for oncology services and episode-based chemotherapy payments that incentivize high quality care. The awardee also planned to continue discussions with commercial payers about their shared savings model, and was exploring a stronger collaboration with its Medicare and commercial accountable care organizations (ACOs) as a way to sustain the clinical components of the program.

Sustainability plans. Cleveland Medical Center made progress in sustaining multiple aspects of its program beyond the cooperative agreement. The awardee reported that some aspects of the program had been organically integrated at Seidman Cancer Center and were expected to continue after the cooperative agreement ended, both at the main campus and the community satellite sites. The awardee also worked on formally integrating other aspects of the program after the cooperative agreement. Awardee leaders also mentioned the possibility of expanding the program's target population after the cooperative agreement ended in order to include people at the medical center with other types of diseases. The awardee planned to simultaneously refine eligibility at the cancer center to ensure the program served those with the greatest need.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a non-clinician staff survey on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July to October of 2016 with a sample of 12 potential respondents and achieved a response rate of 73 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a] Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

Individuals that met the program's eligibility criteria (Table I.1) were primarily identified in two ways. Most often, the care coordination team reviewed reports generated by an internal data management system that contained information on patients' insurance and disease stage. If the disease stage was not included in the report, the nurse care coordinator would confirm the stage with the disease team to confirm the patient's eligibility. The care coordinator would then contact the patient to introduce the program. Alternatively, oncologists at Seidman Cancer Center could refer patients. If an oncologist believed a patient met the eligibility criteria and could benefit from the program, the oncologist could reach out to the LINCC nurse care coordinator and ask the coordinator to meet with the patient. The first method of identifying participants through the data management system was implemented at the beginning of the

second year and substantially reduced the time spent identifying eligible patients, which rapidly increased the rate of enrollment without affecting the eligibility criteria.

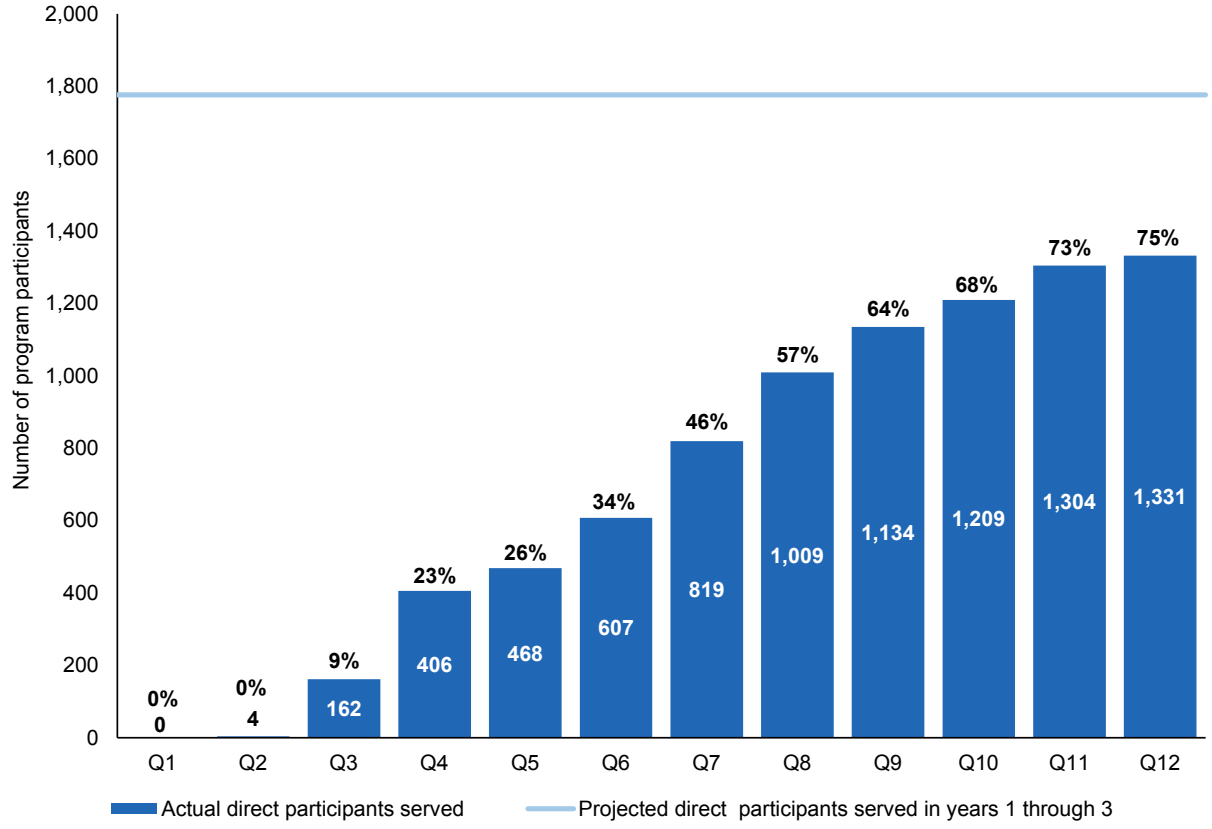
Nurse care coordinators reported that during the third program year, they focused their recruiting efforts on patients whom they believed would benefit most from the intervention. Their focus evolved from enrolling as many patients as possible to taking a more active, targeted approach. The nurse care coordinators used chart reviews, physician notes, frequency of phone calls between patients and the care team, and face-to-face interactions with the disease team to determine which patients would be best served by the program. For example, the nurse care coordinators might look for factors associated with poor outcomes (for example, triple negative breast cancer),[4] poor social support, and need for active symptom management.

b. Evidence of enrollment effectiveness

Overall, the awardee reported that it enrolled 1,331 direct participants from February 2015 (when it launched its program) through August 2017, which represents about 75 percent of its final three-year projections (Figure II.1). The awardee primarily enrolled Medicare and Medicaid beneficiaries, and also enrolled a small number of privately insured patients. The latter are outside the scope of our evaluation. Qualitative evidence suggests that the eligibility criteria are applied consistently across implementing sites; the third eligibility criterion for "complex patients" (that is, patients with other risk factors for poor outcomes and higher spending levels) is somewhat at the discretion of referring physicians and LINCC staff, although LINCC staff thought this category represents a small percentage of the enrolled participants.

[4] About 10 to 20 percent of breast cancers have cells that test negative for estrogen, progesterone, and HER2 receptors, and are therefore called triple-negative. Triple-negative breast cancers do not respond to hormonal therapy or therapies that target HER2 receptors, although there are other therapies available. Triple-negative breast cancers tend to be more aggressive than other types of breast cancer.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding.

c. Barriers and facilitators associated with enrollment effectiveness

Cleveland Medical Center's progress in meeting its three-year enrollment goals was influenced by several factors. At the beginning of the second program year, the awardee rapidly increased enrollment by transitioning to an automated process for identifying patients for the LINCC program. During the second program year, Cleveland Medical Center also expanded the number of care coordinators from three (as originally planned) to seven, which allowed the program to expand to the community satellite locations and thus increase the pool of potential participants. According to program staff, better communication between the LINCC and disease teams, as well as increased recognition by the disease teams of the value of the program's care coordination and palliative care services, sped up the pace of enrollment during the third program year. For example, providers were more likely to refer patients to the program if they viewed the nurse care coordinator as a team member who could take care of a patient's specific needs. Similarly, one nurse care coordinator talked about how the oncologists came to recognize the palliative care providers' skill set and how the growing trust between the oncologist and

palliative care providers helped the program get more referrals from staff external to the LINCC team.

Cleveland Medical Center leaders believe that the inability to fully staff the LINCC program at the community satellite locations and expand to two of the originally planned satellite locations limited enrollment. Awardee leaders described how even though they moved nurse care coordinators from the main campus to the satellites to care for patients, the satellite sites were never fully staffed at the capacity that was originally expected. The awardee reported that the lack of capacity at the satellite clinics forced it to focus its resources on those patients who might benefit most from the intervention.

The accessibility and completeness of health IT also influenced the program's ability to identify and enroll patients. For example, the automated list of eligible patients only contained the patient's disease stage if it was documented properly in the electronic medical record. Some LINCC staff thought that lack of disease stage documentation made it hard to identify eligible patients. However, nurse care coordinators and palliative care providers were able to overcome this by communicating more often with the oncologists and the disease teams. Lastly, some of the staff who participated in the non-clinician staff survey speculated that potential participants may have declined enrollment because they perceived that the technology components (that is, the iPad assessments) were too complicated (patients were not considered enrolled until they had completed the iPad assessments).

2. Delivery of program services

a. Description of and changes to service delivery model

The LINCC program was an entirely new program at the University Hospitals Seidman Cancer Center. Between September 2014, when the cooperative agreement was awarded, and February 2015, when the first participant was enrolled, the awardee (1) recruited and onboarded key personnel including two palliative care providers, two nurse care coordinators, a program analyst, and an administrative assistant; (2) secured internal administrative space; (3) developed an electronic patient assessment; and (4) collaborated with the hospital IT team to enhance the electronic medical record. The program was operationalized within a large academic hospital system, and the roughly 150 providers who practice medical, radiation, surgical, or community oncology needed to be made aware of the program. Moreover, the LINCC program services had to be integrated into the existing busy clinical workflow.

The LINCC program's theory of action began with participant identification and recruitment by the nurse care coordinators or other hospital staff. The nurse care coordinators administered the patient assessment to identify the patient's physical, emotional, and spiritual needs, which they communicated to the disease team, palliative care providers, and spiritual care providers as needed. The nurse care coordinators worked with the patients to develop a plan of care, and connected them to different resources (for example, psychiatry, social work, palliative care) depending on their needs. The palliative care providers met with patients to address their symptoms and engage in discussions about their goals and values. Nurse care coordinators helped patients manage their care and priorities by checking in regularly with them, answering their phone calls, coordinating their various clinical and non-clinical appointments, and helping them adhere to the plan of care. As a result, the patients improved their self-efficacy, and disease

teams adhered to evidence-based practices consistent with the plans of care. This increase in evidence-conformant care was expected to reduce utilization of chemotherapy and of intensive and emergency care, any of which may only intensify suffering in the last days of life; increase the number of patients with pain care[5] and advance care plans; increase use of outpatient clinics; and increase the number of patients admitted to hospice.

Several minor adjustments were made to the initial operational plan to optimize the program's effectiveness and promote its sustainability.

First, modifications were made to the administration mode and frequency of the program's biopsychosocial assessments. Initially, an extended version of the patient assessment was given to all participants. As of February 2016, the extended assessment was only given to enough participants to reach the minimum number required for Cleveland Medical Center's internal quality-of-life analysis. The rest of the participants were given an abbreviated version. In the seventh program quarter, the awardee developed a protocol for administering the assessment over the phone. In the eighth program quarter, the awardee added a new tool to assess spiritual needs as well as a "face sheet" that summarizes screening domains. As of the 12th quarter, Seidman Cancer Center planned to pilot a single streamlined instrument (based on the LINCC instrument and electronic platform) that combines elements of assessments required for other accreditation processes (that is, those of the Joint Commission and the Quality Oncology Practice Initiative) at the end of calendar year 2017.

Second, although in the original operational plan the awardee proposed contacting patients every month, program leaders thought they could customize the program to better reflect patients' needs and the disease's acuity, and thus they adapted the operational plan to require contacting patients halfway between scheduled appointments instead (and not more often than monthly). This limited burden on less symptomatic patients.

Third, during the third program year, palliative care providers started taking care of LINCC participants in both inpatient and outpatient settings, which improved continuity of care and gave patients and Seidman Cancer Center staff more exposure to the LINCC team. Finally, the awardee determined that an optimal caseload for the nurse care coordinator was about 100 patients, not the 250 estimated in the original operational plan, and that the exact caseload should vary with the patients' disease type and acuity.

b. Evidence of service delivery effectiveness

Cleveland Medical Center was successful in achieving implementation effectiveness. Specifically, the program successfully provided program services and engaged participants and providers, but struggled to recruit and retain key staff and experienced delays implementing the program at the community satellite locations. We provide details below.

[5] The program removed "number of patients with pain care" as an outcome in quarter 13 of the program.

Delivery of intervention services. The LINCC program successfully delivered care coordination, palliative care services, and biopsychosocial assessments, but experienced some delays in full implementation of the program.

First, interviews with nurse care coordinators and the nurse manager revealed that LINCC staff delivered the intended care coordination services, such as identifying participants' physical and emotional needs, connecting participants to clinical and non-clinical resources that reflect those needs, helping participants navigate their various appointments and tests, serving as a resource when participants had questions, helping participants establish goals for their care, and helping patients make more informed decisions about their care. Staff responding to our survey corroborated these findings, saying their job included recruiting and enrolling participants; educating participants about managing their own care and services; calling participants to check on medications, compliance, and symptoms or to help coordinate care; and following up on transitions of care and services. Awardee leaders also tracked the caseloads of nurse care coordinators and verified that caseloads ranged between 70 and 150 participants, depending on the disease type and patient acuity.

Second, the LINCC program increased access to and utilization of palliative care at both the main campus and community satellite clinics. According to LINCC self-monitoring measures from the 12th quarter, 70 percent of all participants served to date had received at least one palliative care consult. One hundred percent of participants enrolled during the 12th program quarter had received a palliative care consult. The awardee also reported that for almost all program quarters, palliative care appointments were available on the same day the patient enrolled in the LINCC program, indicating sustained access to palliative care.

Third, the awardee successfully deployed the biopsychosocial assessments. Every LINCC program participant received at least the initial screening, and by the end of the third program year, nearly all patients receiving care at Seidman Cancer Center were receiving follow-up assessments.[6] Nurse care coordinators and palliative care providers used the results of the assessments to inform care coordination activities and palliative care appointments. However, the awardee reported that some patients who received the initial assessment did not take the next steps to complete the assessment, diminishing the effectiveness of the tool for those patients.

Lastly, the awardee experienced delays in establishing the primary program components at the community satellite locations. As of the fifth quarter, one palliative care provider was serving as both palliative care provider and care coordinator at two of the community satellite clinics. During the seventh quarter, nurse care coordinators were deployed to the community satellite sites, and the program was ultimately implemented at four of the six planned sites. Furthermore, the LINCC spiritual care coordinator was onboarded during the sixth program quarter, at which time the program began providing inpatient and outpatient spiritual support to participants. Before this position was filled, participants were able to access spiritual care through Cleveland

[6] These assessments, though based on the LINCC assessments and the electronic platform, were different than the ones used earlier by LINCC. These assessments consisted of a single streamlined instrument that integrated the different earlier screening instruments on patient experience. The single instrument was developed in response to frustration on the part of patients and staff with multiple instruments asking similar questions.

Medical Center chaplains. Finally, the awardee did not hire a pharmacist for the program as originally intended, but it was able to secure time from a medical oncologist. The medical oncologist worked to improve use of appropriate therapies for patients with the PD-L1 biomarker[7] and to standardize the use of anti-emetic drugs.

Staffing and training. Cleveland Medical Center struggled to recruit and retain the planned number of nurse care coordinators to implement the program at its projected capacity. At the end of the first program year, the LINCC program had three nurse care coordinators. Awardee leaders modified the operational plan to include nine positions for nurse care coordinators. However, the maximum number of nurse care coordinators the program had was seven during the second program year. This dropped to five nurse care coordinators in the first three quarters of the third year and three nurse care coordinators as of the 12th program quarter. The awardee was below its total staffing projections at the end of each 12-month period (Table II.2).

Table II.2. Overview of awardee staffing, by end of 12-month period

	Q4	Q8	Q12
Total staff FTEs	10 FTE	12 FTE	8.85 FTE
Total nurse care coordinators	3	6	3
Staffing as a percentage of target	Below by 24%	Below by 35%	Below by 55%

Source: Quarterly reports submitted to the implementation and monitoring contractor.
FTE = Full time equivalent.

Engagement of providers. Program leaders and staff reported that providers were more or less engaged depending on what disease team they were on, but overall, providers grew more engaged with the program over the course of the award. For example, the LINCC leadership team administered an internal survey during the 10th quarter to assess providers' engagement and incorporate their feedback in plans to sustain the program. According to the awardee, most (85 percent) of the physician respondents thought the program improved the quality of care and enhanced physicians' interactions with patients.[8] These physicians thought the program components with the most value were the early assessments of patients, early access to palliative care, and ongoing psychosocial support from the coordinators.

During site visits at the end of the third program year, nurse care coordinators reported that more referrals from physicians on the disease team and greater familiarity with the LINCC program among Seidman Cancer Center disease teams were an indication that providers had become more engaged in the program. Nevertheless, nearly all respondents to the staff survey (fielded by the evaluation team between July and October 2016) indicated that physicians' resistance to the program was either a minor or major barrier to achieving program goals.

Engagement of program participants. In reviewing the survey responses from non-clinician staff, we found that the majority of LINCC staff thought the program had successfully

[7] The PD-L1 test suggests that a tumor may respond to a new class of cancer drugs known as immune checkpoint inhibitors; however, these drugs are still being researched and are extremely expensive.

[8] Data were provided by the awardee; the sample size and response rate are unknown.

engaged participants. Similarly, staff interviewed at the end of the third program year agreed that the nurse care coordinators engaged patients by connecting them with supportive services such as dieticians, counseling, psychiatry, music therapy, art therapy, spiritual care, and pain management. This extra level of care was appreciated by the patients, and the staff thought this helped the patients stay engaged. The palliative care providers believed the program did not end up seeing as many patients for *proactive* palliative care as initially intended, because it was hard to engage asymptomatic patients and their providers. However, as of the end of the third program year, the majority of program participants had received a palliative care consultation.

c. Barriers and facilitators associated with service delivery effectiveness

Over the three-year cooperative agreement, facilitators of service delivery effectiveness were successful teamwork, communication among the LINCC team and between LINCC staff and the disease teams, skilled nurse care coordinators, and the ability to communicate the value of the services to patients and referring providers. Barriers to service delivery included challenges integrating the program components into existing process, staffing shortages, and the limited duration (three years) of the cooperative agreement.

> "You're there to help ease their [oncologists'] load so they can focus more on the treatment of the cancer, and you can focus more on all the symptoms."
>
> —Nurse care coordinator

LINCC staff said that effective communication and teamwork were essential for engaging Seidman Cancer Center disease teams. For example, nurse care coordinators thought they engaged providers most effectively when they confidently reached out to the disease team oncologist about shared patients, and asked the oncologist upfront about his or her preferred mode of communication. Nurse care coordinators also engaged providers by emphasizing how they could work with the providers to give patients the best care and by assuring the disease teams that they were not taking their patients away. Over the course of the three-year cooperative agreement, there were varying levels of engagement from the different disease teams, but effective messaging and evidence of positive results increased providers' acceptance of palliative care and of the program as a whole.

Patient engagement was facilitated by (1) participants' understanding the value of the program, (2) LINCC staff's explanation of palliative care in a way that made it acceptable to participants, and (3) making participation as convenient as possible for participants. For example, the nurse care coordinators and spiritual care coordinator found that patients were more open to receiving palliative care if the staff referred to it as "symptom management," because many patients associated palliative care with hospice and end-of-life care. The program also carefully monitored patients who initially declined palliative care so that as soon as they had symptoms, they were able to set up a visit with a palliative care provider. Staff also made program services more accessible by providing coordination over the phone, meeting patients on days the patients were already in the clinic, and providing services at the community satellite locations, which

> "Let the oncologist focus on the actual treatment and making sure that that stays in place and all the testing that needs [to be] done and all of the scans that need [to be] done [are done], and palliative medicine can handle all of the symptoms associated with that cancer. When it's explained like that [to participants], people have buy-in, and they understand."
>
> —Nurse care coordinator

are outside of downtown Cleveland and often more comfortable and convenient for patients. Barriers to participant engagement included technological challenges (for example, with the patient assessments on the iPads) and asymptomatic patients, who were not interested in care coordination services or palliative care.

During the first year, staff encountered challenges integrating the program services into the existing clinical workflow and physical infrastructure. Clinical integration improved over time because LINCC staff were flexible and creative about where and when they met with patients to minimize disruptions. Staff also administered patient assessments over the phone, demonstrated the value of program services to providers (through both anecdotes and data), and worked toward sustainability of certain program components so disease teams were less hesitant to incorporate them into the workflow. Nevertheless, as of the 12th program quarter, Cleveland Medical Center was still reporting that "changing culture and securing a place for the services in the overall model of care" was its most significant challenge. Some possible reasons for this were gaps in clinical informatics, the program's existence within a large and complex organization, and the necessity of having supporting data before integrating innovation into the standard of care. Awardee staff thought conducting the biopsychosocial assessments had negative impacts on patient throughput, and some frontline staff thought the results were not always immediately accessible to the oncologists and palliative care providers. As of the 12th quarter, the awardee was still making updates to the biopsychosocial assessment face sheet to make the information more accessible.

According to LINCC staff, the program's inability to reach staffing projections reflected a regional nursing shortage, the selective process for hiring nurse care coordinators, and concerns about the sustainability of the position at the end of the cooperative agreement. Specifically, the awardee reported that the nursing shortage is not unique to the LINCC program, but extends to the entire northeast Ohio region. There also were a number of interested applicants who were not extended job offers because the awardee did not think those applicants could meet the demands of the position.

To address these staffing challenges, LINCC staff worked with the hospital's staffing agency and began to recruit nurses from hospice, palliative care, and home care settings; these nurses had backgrounds that translated well to the demands of the LINCC program. To dispel fears about job security, LINCC leaders worked with human resources to develop a sustainability plan, but they thought this did not have a noticeable impact in retaining nurse care coordinators and improving their morale. Program staff and leaders thought that despite staff turnover, they had a group of dedicated nurses to see the program through, and that the replacement nurses were so adept that their hiring did not negatively affect the delivery of services. However, about three-quarters of survey respondents said that staff turnover was either a major or minor barrier to achieving program goals.

One final challenge noted by a number of program staff was the three-year time limit on the program. Many staff thought the three-year limit pressured them to quickly hire staff and implement the program. When problems arose, they did not think they had enough time to find a solution. Similarly, some staff thought they were pressured to meet staffing goals to hit enrollment targets, rather than waiting for the person with the right skill set and personality for the job. Several staff members believed the program was hitting its stride in Year 3: it was

operating more smoothly, and they thought the LINCC staff were at last becoming well integrated with the disease teams.

C. Assessment of perceived program effects on the delivery of care and outcomes

The awardee staff interviewed during the third program year believed that the overall program goals of improving the quality of care, improving patient experience of care, and reducing the total cost of care would be attainable for the program. Respondents indicated that the nurse care coordinators' role in providing an extra layer of support and follow-up was the key to the program's success. Awardee leaders reported that although these goals were attainable, they still needed to determine which patients benefit most to ensure continued sustainability of the program. On the other hand, one respondent said that no matter how effective the program's implementation was, it would be tough to change use of the emergency department by end-stage cancer patients.

LINCC staff who participated in the third year of site visits unanimously agreed that the program was most effective in its third year. The awardee also reported to the implementation and monitoring contractor that its own initial internal analysis suggested the program had a favorable impact on gross per beneficiary per month costs, quality of life, and patient satisfaction. As of the last program quarter, the awardee was planning to begin a second impact analysis to confirm the initial results with more statistical power.

Interview respondents gave anecdotal evidence of the ways that care coordination, spiritual care, and palliative care could affect program outcomes. Several respondents noted that participants have an extra layer of support to help keep track of their appointments, manage their symptoms, and make educated decisions about their care. For example, having a nurse care coordinator check in on patients often can decrease the likelihood of a patient's going to the emergency department. Similarly, the nurse care coordinators prevent readmissions by making sure everything is in order at discharge, such as the patient's medications and home health care. Staff interviewed during the third year of site visits also said that greater personalization of care improves the experience of care in four ways: (1) patients have nurses' direct numbers, (2) patients have more information to make decisions, (3) follow-up calls bring care to their home instead of requiring them to come to the hospital, and (4) the spiritual and palliative components of the program bring peace to patients who need to consider hospice.

> "One of the things that I've learned the most from the grant in palliative care is how helpful it is to have a really skilled nurse working closely with you … So if I see someone and they're in a lot of pain, I'll ask the care coordinator to call them in four days or whatever to see how their pain is and [whether it] is … better. I've never worked in a setting where I had staff that was capable that had the time to do that and was capable of doing that and doing it well."
>
> —Palliative care provider

Most respondents to the survey of non-clinician staff thought the program had a positive impact on several domains of care, including the (1) quality of care given to participants, (2) ability to respond quickly to participants' needs, (3) efficiency of care, (4) ability to provide care or services that are responsive to participants' preferences, needs, and values, (5) fair provision of services to all participants, (6) access to care for all participants, (7) participant satisfaction, and (8) quality of life. In addition, the majority of respondents thought the LINCC program had

successfully engaged its participants, collaborated effectively to meet participant needs, and led to better treatment plans and/or outcomes. There were varying opinions on the *overall* effectiveness of the program, with half of survey respondents indicating the program was either very or somewhat effective, and the rest saying they thought it was very or somewhat ineffective, or it was too soon to tell.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

Program leaders and staff generally believed the program had a positive impact on participants. Implications and considerations for the impact evaluation are described below:

- **Timing**: Given minor adjustments to improve the program over time, delayed expansion to the community satellite sites, delays in hiring the spiritual care coordinator, improved clinical integration over time, and greater focus on continuity of care during the third year, we might expect program impacts to be strongest later in the cooperative agreement, especially during the third program year. In addition, the enrollment pace during the first year was slower than anticipated, with 162 participants at the end of the third program quarter and 406 participants at the end of the first program year. The small sample size in the first year made impacts during the first year of implementation difficult to detect.
- **Population size and characteristics.** Although the awardee was able to enroll 1,331 participants by the end of the 12th quarter, the impact analysis currently includes only 410 participants who were fee-for-service (FFS) Medicare or dual enrollees who could be matched with at least one control in the comparison group. The size of the treatment group may limit the statistical power necessary to detect smaller effects. Furthermore, awardee leaders believe that the program effects will be greatest among younger, low-income individuals enrolled in Medicaid. Because we are unable to obtain Ohio Transformed Medicaid Statistical Information System (T-MSIS) data at this time, in the current impact chapter we only focus on participating Medicare beneficiaries. Medicaid data may be included in the future depending on the quality of T-MSIS data and the number of Medicaid enrollees in the program.
- **Diagnosis.** The awardee believes the program impact will vary with the patient's disease type and acuity. Therefore, we might expect to see greatest impacts among people with certain types of cancer or among the enrollees who were sickest at the time of enrollment. As described above, nurse care coordinators focused their enrollment efforts on those most likely to benefit from LINCC services during the third program year, so it is possible that these higher need patients will have more room for improvement. We have yet to determine whether the treatment group will be large enough to conduct subgroup analyses by disease type.
- **Impact by outcome.** Finally, LINCC staff believe that the program will have the greatest impacts on experience and quality of care, and they are less certain about impacts on cost and utilization (or think it is too soon to know). The impact analysis will not evaluate experience of care because there are no outcomes on experience of care available for the

comparison group. The impact evaluation will look at measures associated with the quality of care, such as hospice use, intensive care unit use, and chemotherapy or radiation at the end of life. Decreases in avoidable emergency department visits, hospitalizations, and 30-day hospital readmissions are not among the awardee's primary program goals, but these measures are among CMMI's core outcomes for the HCIA R2 initiative and will be examined as part of this evaluation.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Cleveland Medical Center's LINCC program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: Cleveland Medical Center

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	457[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	178[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	634
Likelihood of all-cause hospitalizations	489
MDE sample size requirement to detect 20% effect	
Total expenditures	159
Likelihood of all-cause hospitalizations	122
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group
Likelihood of solid comparison group	Some issues, but probably surmountable; expect to select a comparison group
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff survey
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We plan to carry out a rigorous analysis of the LINCC program. The sample is projected to be large enough to detect effects of less than 20 percent of Medicare spending. Our initial effort to select a matched comparison group by using data from the Ohio cancer registry program appeared successful at first, producing a comparison sample that was well matched to treatment beneficiaries with respect to age, sex, cancer site, cancer stage, county of residence, and prior Medicare spending. Subsequent investigation, however, showed that the comparison sample did not adequately match the trajectory of spending and utilization for the treatment group in the pre-enrollment period. We are now repeating the matching process in an effort to select a comparison group that better mirrors the treatment group's pre-enrollment spending and utilization.

B. Characteristics of Medicare and Medicaid participants at baseline

For the purposes of this report, the Cleveland Medical Center's treatment group consisted of Medicare FFS beneficiaries who enrolled in the awardee's program between February 19, 2015, and May 31, 2016.[9] In the future, the Medicaid sample is anticipated to be large enough to detect an impact on key outcomes and may be included in the impact evaluation at a later date if the data become available.

In presenting the baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS (both Parts A and B) with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who had met all evaluation criteria for a period of at least 90 days during the baseline year (the 365 days immediately before their enrollment). The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. After we excluded beneficiaries who did not meet the above criteria, a total of 327 participants were included in the analysis of baseline characteristics for this report.

The demographic characteristics of program participants are similar to those of Medicare FFS beneficiaries nationwide (Table III.2). Most participants are either 65 to 74 years old (48 percent) or 75 to 84 years old (29 percent). The sample skews female (55 percent) and is predominately white (73 percent). Fourteen percent of participants are dually eligible for Medicare and Medicaid. The average hierarchical condition categories (HCC) risk score of participants (3.74) is nearly four times higher than the national average for Medicare FFS beneficiaries (approximately 1.00). Nearly all of the participants have HCC risk scores higher than the national average.

[9] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS and those participants for whom the finder file has valid Medicare identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

Program participants had high and increasing rates of service use and Medicare expenditures in the baseline year. Table III.3 shows baseline utilization and expenditure data for a common set of measures. We examined baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $3,719—far above the U.S. average of $792.[10] Average PBPM Medicare payments for acute inpatient ($1,383) and outpatient ($1,371) services were the largest drivers of total cost of care.

[10] See the report on Medicaid enrollment from the Medicare Payment Advisory Commission and the Medicaid and CHIP Payment and Access Commission at http://www.medpac.gov/documents/data-book/january-2015-medpac-and-macpac-data-book-beneficiaries-dually-eligible-for-medicare-and-medicaid.pdf.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Cleveland Medical Center's program through May 31, 2016

	All participants (N = 327)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	46	14
65 to 74	158	48
75 to 84	95	29
85 and older	28	9
Gender		
Female	179	55
Male	148	45
Race		
White	240	73
Black	75	23
American Indian, Alaska Native, Asian/Pacific Island American, or other	5	2
Hispanic	1	0.31
Original reason for Medicare eligibility		
Old age and survivor's insurance	261	80
Disability insurance benefits	66	20
End-stage renal disease (ESRD)[a]		
Hospice[b]	3	0.92
Medicare/Medicaid dual status, percentage dual[b]	46	14
HCC score[c]		Statistic
Mean		3.74
25th percentile		2.46
Median		3.54
75th percentile		4.8

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the first day of the month in which the beneficiary began receiving HCIA R2-funded services. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Cleveland Medical Center's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	327	303	309	326	327
Average Medicare expenditures PBPM[a]					
Total	3,719 (192)	2,645 (285)	2,881 (339)	3,578 (331)	5,581 (343)
Acute inpatient	1,383 (118)	904 (172)	909 (253)	1,219 (194)	2,380 (262)
Inpatient other[b]	61 (23)	57 (41)	29 (26)	92 (50)	66 (33)
Outpatient[c]	1,371 (102)	1,041 (151)	1,144 (128)	1,334 (148)	1,921 (115)
Physician services	617 (47)	449 (48)	541 (63)	602 (90)	857 (50)
Home health	93 (11)	62 (15)	69 (17)	119 (24)	120 (19)
Skilled nursing facility	123 (25)	75 (29)	122 (52)	151 (46)	142 (40)
Hospice	12 (7)	0 (0)	0 (0)	13 (9)	34 (18)
Durable medical equipment	59 (9)	58 (19)	66 (15)	49 (10)	61 (9)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	1,184 (89)	747 (116)	824 (141)	1,040 (149)	2,023 (174)
Outpatient emergency department (ED) visits	795 (119)	520 (109)	720 (131)	723 (130)	1,189 (220)
Observation stays	183 (29)	13 (13)	183 (54)	216 (171)	306 (61)
Primary care visits in any setting	7,685 (389)	6,162 (619)	6,530 (596)	7,193 (663)	10,640 (697)
Primary care visits in ambulatory settings	5,376 (259)	4,601 (347)	4,972 (436)	4,910 (351)	6,913 (421)
Specialist visits in any setting	18,982 (761)	14,724 (1,031)	15,022 (961)	18,510 (1,109)	27,053 (1,143)
Specialist visits in ambulatory settings	14,155 (515)	11,710 (682)	12,026 (640)	14,031 (753)	18,510 (628)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with hospital admission[d]	60 (3)	14 (2)	15 (2)	18 (2)	37 (3)
Percentage with an outpatient ED visit[e]	41 (3)	10 (2)	13 (2)	14 (2)	19 (2)
Percentage with an observation stay[f]	16 (2)	0 (0)	4 (1)	5 (1)	7 (1)
Percentage with a 30-day readmission among all discharges	26 (3)	20 (6)	35 (6)	20 (5)	28 (4)
Percentage of participants with a readmission among all participants	16 (2)	3 (1)	3 (1)	3 (1)	8 (2)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

The rate of acute care hospitalizations for participants was 1,184 per 1,000 Medicare FFS beneficiaries per year during the baseline year (a rate much higher than the U.S. average of 274 per 1,000 Medicare FFS beneficiaries per year[11]), with 60 percent of participants having at least one hospitalization during the baseline year. The rate of acute care hospitalizations for participants was highest in baseline quarter 4 (2,023 per 1,000 Medicare FFS beneficiaries per year) compared with baseline quarters 1 through 3 (747 to 1,040 per 1,000 Medicare FFS beneficiaries per year). About 37 percent of participants had at least one hospitalization during baseline quarter 4.

The rate of emergency department (ED) visits that did not lead to a participant's hospitalization in the baseline year was 795 per 1,000 Medicare FFS beneficiaries per year. The rate of observation bed stays in the baseline year was 183 per 1,000 Medicare FFS beneficiaries per year. The rate of primary care visits in any setting was 7,685 per 1,000 Medicare FFS beneficiaries per year. The rate of specialty visits in any setting was 18,982 per 1,000 Medicare FFS beneficiaries per year. All of these rates were higher in baseline quarter 4 compared with quarters 1 through 3.

In the baseline year, 26 percent of hospital discharges among all discharges were followed by a readmission in the 30-day post-discharge window, whereas 16 percent of all Medicare FFS beneficiaries had a hospitalization with a readmission in the 30-day post-discharge window.

Table III.4 provides estimates of the use of chemotherapy and non-palliative radiation therapy, hospice, and ICU admissions among deceased beneficiaries in the treatment group to date. Of the 113 individuals in the treatment group who died, 20 percent received chemotherapy or non-palliative radiation therapy in the last 14 days of life. About 35 percent of the deceased treatment group patients received no hospice care. In addition, 11 percent of the deceased treatment group patients were admitted to the ICU in the last 30 days of life. Future reports will provide updated estimates of these measures in both the treatment and comparison groups.

Table III.4. Awardee-specific measures for Cleveland Medical Center

Measures	Denominator	Numerator	Percentage
Percentage of deceased patients who received chemotherapy or non-palliative radiation therapy in the last 14 days of life	113	23	20.4
Percentage of deceased patients who were not admitted to hospice	113	39	34.5
Percentage of deceased patients admitted to the ICU in the last 30 days of life	113	12	10.6

Note: Average rates adapted from prior literature: chemoradiation therapy, 8 percent (adding 2 percentage points to account for non-palliative radiation therapy); hospice nonadmission, 46 percent; ICU admission, 25 percent. Source: Morden, N. E., C.-H. Chang, J. O. Jacobson, E. M. Berke, J. P. W. Bynum, K. M. Murray, and D. C. Goodman. "End-of-Life Care for Medicare Beneficiaries with Cancer Is Highly Intensive Overall and Varies Widely." *Health Affairs,* vol. 31, no. 4, 2012, pp. 786–796.

[11] For national average rates, see the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation" at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed July 1, 2016.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Cleveland Medical Center was proceeding with participation in the CMS-sponsored OCM, a program launched by CMS in July 2016 that involves monthly enhanced payments for oncology services and episode-based chemotherapy payments that incentivize high quality care. The awardee also planned to continue discussions with commercial payers about their shared savings model, and was exploring a stronger collaboration with its Medicare and commercial ACOs as a way to sustain the clinical components of the program.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Cleveland Medical Center joined CMS's OCM, was in discussions with University Hospitals' Medicare ACO, and had developed a shared savings model to negotiate with commercial payers.

The OCM involves a $160 PBPM fee on top of regular FFS reimbursements. The fee covers LINCC services for Medicare FFS beneficiaries beginning with the first episode of chemotherapy and continuing for six months. In addition, participating providers are eligible for semiannual performance-based payments for savings compared to a risk-adjusted target amount above 4 percent (one-sided risk and two-sided risk arrangements are available). The portion of shared savings is adjusted by a performance multiplier based on an aggregate quality score constructed from each practice's or pool's performance on 12 quality measures:

- OCM-1 Risk adjusted proportion of patients with all-cause hospital admissions
- OCM-2 Risk-adjusted proportion of patients with all-cause ED visits that did not result in a hospital admission
- OCM-3 Proportion of patients who died who were admitted to hospice for three days or more
- OCM-4 Pain assessment and management
- OCM-5 Preventive care and screening: screening for clinical depression and follow-up
- OCM-6 Patient-reported experience of care
- OCM-7 Prostate cancer: adjuvant hormonal therapy for high-risk beneficiaries
- OCM-8 Timeliness of adjuvant chemotherapy for colon cancer
- OCM-9 Timeliness of combination chemotherapy for hormone receptor negative breast cancer

- OCM-10 Trastuzumab received by patients with AJCC stage I (T1c) to III Her2/neu positive breast cancer
- OCM-11 Hormonal therapy for stage IC-IIIC estrogen receptor/progesterone receptor positive breast cancer
- OCM-12 Documentation of current medication

Many, but not all, of the HCIA R2 participants would be eligible for services under this funding mechanism. Those who would be ineligible include people who are not being treated with chemotherapy, people receiving benefits for treatment of end-stage renal disease, and individuals insured by Medicare Advantage or other group health programs.

If the program is sustained as part of the ACO, the awardee expected that some portion of program operations would be sustained through a direct subsidy from the ACO. The ACO had negotiated a coordination fee and shared savings arrangements with various commercial and government payers.

The shared savings model was the least developed of the awardee's models, because there had been limited success engaging commercial payers in negotiations on that model. The awardee's shared savings model included savings based on spending targets and a $160 PBPM fee that would cover LINCC services with the goal of reducing costs.

C. Status of the payment model

Cleveland Medical Center was proceeding with participation in the OCM. The awardee was exploring a stronger collaboration with the ACO pending the results of the quantitative evaluation. Cleveland Medical Center continued discussions with commercial payers about its shared savings model, but the negotiations had not yielded any agreements as of October 2017.

D. Factors associated with the development of the payment model

The awardee cited three major challenges with developing payment models for the program: (1) lack of data, (2) payer hesitation with condition-specific payment models, and (3) a lack of evidence for the payment models. First, the lack of data on whether the program saves money was identified as a key hindrance. Second, the awardee reported that "payers communicate their hesitation to implement a model that represents (another) carve out to their existing contractual arrangements and billing/actuarial systems." Third, the awardee stated that the policy environment has been a challenge, because there is no clarity on what payment mechanisms and incentive/penalty arrangements work best to lower costs while improving care quality. On the positive side, the awardee cited senior executive support for the program.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Cleveland Medical Center made progress in sustaining multiple aspects of its program beyond the cooperative agreement. The awardee reported that some aspects of the program had been organically integrated at Seidman Cancer Center and were expected to continue after the cooperative agreement ended, both at the main campus and the community satellite sites. The awardee also worked on formally integrating other aspects of the program after the cooperative agreement. Awardee leaders also mentioned the possibility of expanding the program's target population after the cooperative agreement ended in order to include people at the medical center with other types of diseases. The awardee planned to simultaneously refine eligibility at the cancer center to ensure the program served those with the greatest need.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Cleveland Medical Center planned to sustain some aspects of its program. Specifically, the awardee planned to sustain the biopsychosocial screening tool, but was waiting for program data to understand the impact of care management provided by nurse care coordinators before creating a sustainment plan for care coordination. Cleveland

Medical Center had also expanded access to palliative care and the biopsychosocial assessments, which were offered to patients beyond the program's original target population at both the main campus and community satellite clinics. The medical center expected to sustain this expanded access at the end of the cooperative agreement.

C. Implementing the SSR plan: progress and changes

Sustainability. As of the third program year, Cleveland Medical Center reported plans to sustain multiple aspects of the program. Some aspects of the program were already "organically integrated" by the end of the third program year, meaning hospital staff (beyond the LINCC program) were regularly using program resources without being encouraged to by the awardee. Examples include increased adoption of expert-level palliative care services, regular use of biopsychosocial assessments, and enhancements to the electronic medical record to document advance care planning and goals of care. With regard to the biopsychosocial assessment, the awardee reported that the Seidman Cancer Center will pilot a single streamlined instrument that integrates the different screening instruments on patient experience. This is being done in response to frustration on the part of patients and staff with multiple instruments asking similar questions. The new instrument is based on the LINCC instrument and electronic platform and satisfies existing accreditation and other requirements. It is expected to be rolled out system wide in mid-2018.

Cleveland Medical Center also described plans to formally sustain aspects of the program that had not been organically integrated into sites. Specifically, the awardee reported efforts to sustain the position of nurse care coordinator, which would have higher operational costs and greater impact on clinic operations compared to other components of the program. The awardee participated in discussions with the hospital's ACO about using care coordination payments to subsidize care coordination activities going forward; however, these details were still being discussed, and the outcome would depend on whether data would demonstrate the program's value. The awardee planned to have the business proposal for sustaining care coordination finalized in 2018. In the meantime, the awardee built two full-time equivalent positions for care coordinators into the fiscal year 2018 budget to sustain the program and maintain its momentum. In an effort to optimize the nurse care coordinators' time, awardee leaders discussed using outcome data from patients with complex cancers to determine which patients benefited most from nurse care coordination.

Scalability. Cleveland Medical Center reported plans to possibly scale the program after the cooperative agreement ended to people at University Hospitals with other diseases or conditions, such as bone marrow transplants, heart failure, and solid organ transplants. Awardee leaders also expected that expanded access to palliative care and biopsychosocial assessments would continue for all Seidman Cancer Center patients as a result of the groundwork established by the LINCC program.

Replicability. Cleveland Medical Center did not report plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

Cleveland Medical Center reported that support from organizational leaders helped the awardee make progress in sustaining the program after the cooperative agreement ended. The

awardee reported that organizational leaders supported the program and expressed a desire to sustain it, perhaps because of the LINCC's alignment with the organization's work to obtain certification from the Quality Oncology Practice Initiative, the organization's participation in the Oncology Care Model, and leaders' commitment to delivering value-based care.

However, the awardee also said certain things could keep leaders from continuing to support the program. These included an inability on the part of the awardee to (1) prove the program's value, (2) recruit qualified staff, and (3) predict the political landscape.

First, the awardee reported that it will have to continue demonstrating the program's value to sustain long-term engagement from Cleveland Medical Center's leaders. In particular, the awardee suspected that the leaders' engagement may suffer if the awardee cannot also engage payers; but before that can happen, the awardee will need more data to show the program's value to payers.

Second, the awardee said it was difficult to recruit qualified staff throughout the cooperative agreement. The awardee expected staffing to also challenge program sustainment, especially given the regional shortages of nurses and the need to attract staff whose training and attitudes make them the right cultural fit. Cultural fit is not just an issue with new hires, but also with existing staff who may resist the cultural shift within the organization. The awardee believed that upstream support from leaders helped spread awareness of the program's services and the expected cultural attitudes, and continued support of that kind would be necessary to sustain the cultural shift in the organization.

Lastly, the awardee said it was difficult to know how the political landscape might impact the sustainability of the program and leadership engagement.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Cleveland Medical Center's LINCC program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicare beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Board of Trustees of the University of Illinois, Chicago

April 18, 2018

Jennifer Lyons (Mathematica Policy Research)
Margaret H. Coit (Mathematica Policy Research)
Kate Stewart (Mathematica Policy Research)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

This page has been left blank for double-sided copying.

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group or show balance between treatment and control groups for any randomized controlled trials, and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The University of Illinois received a 12-month extension ending August 31, 2018. It plans to end program enrollment on February 28, 2018, but will continue serving CHECK participants until the end of the no-cost extension period.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

The University of Illinois used funding from HCIA R2 to implement the CHECK program, which was intended to improve care coordination for children and young adults (ages 25 and younger) who have chronic medical conditions and who live in Cook County, Illinois. The program focused on improving outcomes for children with complex medical conditions by addressing both their health care needs (for example, by helping beneficiaries connect with community-based providers), and their social needs (for example, by referring them to social service agencies that can help resolve their problems with housing, food security, and school).

Community health workers did the following: (1) referred families to appropriate social service agencies; (2) coordinated with CHECK mental health staff to address behavioral health needs of the children or their family members (needs were identified during the assessment process); and (3) educated families so they could better manage the youth's health condition(s), and (4) coordinated as necessary with health care providers.

The program was housed in the University of Illinois College of Medicine, but the awardee worked to engage participants "where they are," emphasizing the role of long-term relationships with the participants. The program sent community health workers to meet with participants' families in their homes, social service agencies, community- and school-based health centers, and other local sites convenient for the participants and their families. The care coordination staff sustained relationships with participants in their caseload through secure two-way text messaging, telephone contact, and in-person meetings, with the goal of contact at least every six months. The program's mental health staff provided initial counseling, education, and referrals via in-person sessions, group sessions, and secure videoconferencing platforms.

The awardee received a 12-month no-cost extension ending August 31, 2018. The awardee will stop enrolling participants after February 2018, but will keep providing services until the end of the no-cost extension. Table I.1 provides an overview of the CHECK program.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The Board of Trustees of the University of Illinois, Chicago (UIC) implemented the CHECK program to improve care coordination for children with complex medical conditions in Cook County, IL.
Major innovation	The CHECK program deployed community health workers from the same neighborhoods as the children and families it serves. These staff built relationships with children and their families, identified their needs for health and social service care coordination, and connected participants to information and services to address these needs. The CHECK program's health technology provided resources and communication support and allowed for remote access to mental health services.
Program components	Enhanced care coordination to address medical and non-medical needs, i.e., social determinants of health Mental health services and referrals Development and use of health technology (videoconferencing platform, online self-education portal, two-way text messaging, and care coordination software with integrated social service resource repository) to support care delivery components

Table I.1 *(continued)*

Program characteristic	Description
Target population	Children and young adults (ages 25 and younger) with chronic medical conditions, Participants met the following criteria: • Premature birth or diagnosis of diabetes, sickle cell disease, or asthma • Enrolled either in a Medicaid managed care organization under contract to the CHECK program or in Medicaid FFS
Theory of change/ theory of action	The awardee hypothesized that better access to social services and to primary, specialty, and mental health care will result in better health and social outcomes, including fewer hospitalizations and ED visits, and lower costs. The program implemented enhanced care coordination to address the medical and non-medical needs of children and young adults with chronic medical conditions, and paired it with a program that delivered mental health services and referrals. Community health workers coordinated medical and non-medical services, as well as the delivery of mental health services, with support and guidance from care coordinators who led each care coordination team. The CHECK program implemented new software and consumer-facing technology to support its care coordination activities.
Payment model	A per beneficiary per month (PBPM) care coordination fee paid by Medicaid managed care organizations
Award amount	$19,581,403
Effective launch date	December 2014
Program settings	Community health workers engaged participants and families over the phone, as well as at participants' homes or in other community settings, including school- and community-based health centers. Staff on the mental health promotion team provided direct promotional and early intervention services to participants by phone and in the CHECK offices. Care coordination and mental health services were also provided through health technology (for example, SMS platform and videoconference).
Market area	Urban
Market location	Cook County, IL
Target outcomes	Increase the number of participants and families actively engaged in their own care Improve participants' health and quality of life, including improving school attendance Reduce total cost of care for the patient population

FFS = fee-for-service; ED = emergency department; SMS = short message service.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was partly successful in implementing its program over the course of the initial three-year cooperative agreement.[4] We based this conclusion on five factors. First, the awardee enrolled 18,028 participants, 103 percent of its target. Second, the awardee struggled to develop its service delivery model and related team structure and protocols during the first year of the program. In addition, staff time was absorbed by the need to conduct outreach to and engage new participants, which prevented staff from focusing on coordinating care and promoting participants' mental health. After restructuring the staffing models for care coordination and mental health promotion with the support of new leaders during the second year, the program was ultimately able to consistently deliver intervention services. Third, after

[4] Although the impact evaluation focuses on only the second two years of the CHECK program, our implementation effectiveness assessment evaluates implementation over the course of the full three-year cooperative agreement.

some early delays and challenges, the awardee succeeded in hiring and training staff for its care coordination and mental health promotion functions and, under the revised staffing models, achieved strong rates of retention. Fourth, despite significant challenges finding and engaging enrolled participants, the awardee appeared to have successfully engaged participants in CHECK services by the third year of the program, in part due to the strong fit of the program to participants' social needs. The awardee provided direct services to 8,455 of these enrollees—103 percent of its target—by the end of the initial cooperative agreement.[5] Finally, participating clinicians and other implementation staff reported that the program had a positive effect on the delivery of care.

Impact evaluation. We collaborated with the awardee to randomize beneficiaries from a single Medicaid managed care plan to either the CHECK program or a control group, beginning in the second year of the cooperative agreement. We found a decreased number of outpatient ED visits among CHECK participants compared to control individuals in the first 12 months of this randomized controlled trial (RCT). Most of the decrease took place in the first six months after randomization. In the medium- and high-risk subgroup of participants, CHECK reduced the number of hospitalizations, spending on hospitalizations, and total spending in months 7–12 after randomization.

Payment model. At the end of the third year of its cooperative agreement, the awardee's proposal for a primary payment model was a per beneficiary per month (PBPM) care coordination fee for Medicaid managed care organizations (MCOs), but direct negotiations were on hold pending the re-bidding of MCO contracts by the state Medicaid agency in the fall of 2017. The awardee was also exploring a direct employment relationship with its home health system, but the system had minimal interest in a program focused on children.

Sustainability plans. The awardee reported it was working to analyze program data, both to build a business case for the program that would engage payers and organizational leaders and to determine the most efficient staffing models and operational budgets. The awardee also reported on scaling activities for the third program year, including pilot programs in community-based health centers and an expanded presence in the University of Illinois hospital. The awardee reported that the biggest challenges to sustainment and scaling were the inadequate and unreliable funding sources for care coordination programs for children. University of Illinois worked to overcome this challenge by relying on more than one funding source.

[5] In November 2016, the awardee reported that its final three-year projection for direct participants was 23,712—a significant increase over earlier projections. In May 2017, the awardee submitted a correction to its final three-year projection, noting that the previous projection was based on a computational error. The awardee's final three-year projection was 8,200 direct participants served.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of non-clinician staff on their perceptions of the program's effect on the delivery of care. The staff survey was fielded from July to October of 2016 with a sample of 46 potential respondents, and achieved a response rate of 78 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents. We did not field a clinician survey with this awardee because the program did not directly engage clinicians.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a] Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partly successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. We assessed program implementation as partly successful because the University of Illinois faced significant challenges and delays implementing its core services of coordinating care and promoting participants' mental health during the first two years of the program. However, the awardee ultimately achieved its enrollment and engagement goals, hired and trained staff, and appeared to be providing services consistently during the final year of the program. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

The CHECK program had a passive enrollment strategy, in which CHECK program staff identified eligible participants through Medicaid claims data supplied by partnered MCOs or health centers. The awardee considered participants to be enrolled in the program once an enrollment letter was sent to the family. The awardee did not expect all enrolled participants to receive services from the CHECK program; instead, the enrolled participants served as a base

that CHECK staff would reach out to and engage, further described in the "participant engagement" section below.

The University of Illinois approach to enrollment evolved over the course of the award period to add new sources of eligible participants. In addition to the aforementioned partnerships with Medicaid MCOs and the state Medicaid agency, the awardee also accepted direct referrals to the program from providers within community health centers and the university's health system. Site visit interviews revealed that these direct referrals were rare during the first year of the program, but increased somewhat as the awardee built new relationships in the community and the program became better known within the health system and at community- and school-based health centers. Despite this, program leaders reported that direct referrals still made up a small proportion of overall enrollment during the award period.

Starting in the program's second year, community health workers were assigned to the health centers and built relationships with providers and staff to encourage referrals to CHECK. In the third year of the cooperative agreement, CHECK program leaders reported that they were operating at capacity in terms of engaged participants, and as a result, the awardee ended its use of claims data to identify and enroll new participants in the program except when it was directly requested to do so by a managed care plan. The awardee continued to accept new participants who were referred directly to the program, however.

The awardee's approach to its eligibility criteria also changed slightly to mirror the source of enrollment. In interviews, staff noted that as direct referrals to the program increased, providers occasionally referred patients who were in need of CHECK's services but did not meet the program's diagnosis-based eligibility criteria. Program leaders noted that these referrals were reviewed on a case-by-case basis by clinical leaders to determine the youth's suitability for CHECK. Several program leaders and frontline staff noted that in the future, the program may loosen or revise its eligibility criteria to be more inclusive.

b. Evidence of enrollment effectiveness

Overall, the awardee reported that enrolled 18,028 direct participants in the CHECK program from December 2014 (when it launched its program) through August 2017, which represents about 103 percent of its final three-year projections.[6] We therefore find that the University of Illinois was effective in meeting its enrollment goals.

The awardee revised its projected enrollment numbers significantly upward over the course of the program, from 6,000 to nearly 17,500 enrolled participants. Program leaders reported that they increased their enrollment goal because they faced challenges in contacting potential participants' families and required higher numbers of enrolled participants to meet their goals for the number of participants who directly engaged with the program.

[6] We were unable to provide a figure depicting the number of participants enrolled over the course of the program because these data were not available from the implementation and monitoring contractor. We instead provided the target and actual number of participants enrolled in the program at the end of the third year of the cooperative agreement, which we confirmed with CHECK leaders by email.

c. Barriers and facilitators associated with enrollment effectiveness

Early in the award period, the University of Illinois faced challenges with enrolling participants because only limited data were available to identify eligible children and adolescents. When the University of Illinois applied for HCIA R2 funding, it expected to contract with the state Medicaid agency directly to obtain claims data to identify potentially eligible beneficiaries. Before enrollment began, however, the state transitioned to mandatory managed care for Medicaid beneficiaries. The awardee was required to contract with individual managed care plans to obtain these data, and it received more limited claims data for enrollment than it initially expected to. As a result, enrollment was limited to children and adolescents enrolled in a Medicaid managed care plan that the CHECK program contracted with, or those not yet enrolled in a managed care plan. Data from the managed care plan data often did not show a complete claims history, making it more difficult to identify eligible participants. As a result, enrollment did not ramp up at the pace it was expected to during the first year of the program.

2. Delivery of program services

a. Description of and changes to service delivery model

The CHECK service delivery model had several components that evolved over the course of the cooperative agreement:

- **Care coordination.** To coordinate services to address participants' health and social needs, the University of Illinois hired 31 full-time paid community health workers with experience in community health. Many of them lived in the communities they worked in; many also had some personal connection familiarizing them with the program's targeted health conditions.

  For participants, community health workers were the face of CHECK, and their telephone outreach and in-person visits represented the bulk of participants' direct interaction with the program. Each community health worker was assigned to a care coordination team, which was led by a care coordinator. Care coordinators guided and supported the teams of community health workers. These program staff facilitated team collaboration, assisted with especially complex cases, and troubleshot difficult situations involving participants. The CHECK program had six full-time care coordinators. Most of them were not clinically trained, but had bachelor's degrees in related fields.

  Care coordination teams were assigned to geographic zones and neighborhoods in the program's catchment area, allowing community health workers to become familiar with the needs and concerns of those smaller communities. During the second year of the cooperative agreement, the director of care coordination redesigned the staffing model to include more teams, with fewer community health workers working with each supervising care coordinator. Several of the more experienced community health workers were promoted to care coordinator. Although there was no standardized caseload size for these teams, staff reported in interviews that most community health workers were assigned about 175 participants.

  After an initial assessment of the participants to identify their health and psychosocial needs, community health workers worked with participants and their families to develop a care plan. The CHECK program staff assigned participants to a risk tier (high, medium, or low) based on the number of ED visits and hospitalizations they had in the previous 12 months, as

well as their mental health status as determined in the assessment process. These risk tiers were intended to help care coordination staff target their efforts to participants who are most in need of program support.

Participants received differing program services according to their specific needs. Some participants required significant outreach—for example, home visits from the program's community health workers—and ongoing support to coordinate primary and specialty care appointments; others benefited from occasional telephone calls from community health workers or engagement with the health technology described below.

Community health workers met with participants at their homes, at social service agencies, in the hospital, at community- and school-based health centers, and at other local sites that were convenient for the participants and their families. CHECK community health workers were also assigned to attend clinical rounds in the pediatric units of the university hospital to reach out to CHECK participants who had been admitted. By the third year, the awardee had created a separate, dedicated team of community health workers focused solely on going out into the community to "find" participants who were difficult to reach or had fallen out of touch with the program.

To support condition-specific coordination, the University of Illinois developed care coordination protocols for managing the four targeted conditions (asthma, diabetes, prematurity, and sickle cell disease); oral disease, and mental health. Community health workers who received training on a specific targeted condition specialized in providing coordination and support to participants with that condition.

The awardee also established partnerships with specific service organizations to meet the particular needs of participants and their families. For example, the awardee partnered with a mobile oral health service to provide dental care to participants and their families. In addition, in the third year of the program, CHECK staff collaborated with a medical-legal partnership to help participants and their families address legal challenges such as eviction.

- **Mental health services.** Staff on the team promoting participants' mental health delivered a range of services corresponding to needs identified when the participants were assessed. The structure of the team and its services evolved over the course of the award period. The team was originally composed of two part-time master's level mental health counselors and two part-time externs from the University of Illinois clinical psychology program. During the first year of the program, the mental health promotion team mostly focused on educating participants and families (for example, through the use of informational DVDs distributed to parents). In the second year of the program, the director of this team left the program, and one of the master's level counselors was promoted to the director position. The new director redesigned the team structure, shifting from several part-time staff to a smaller core team of two dedicated full-time staff. This enhanced their ability to conduct regular mental health assessments, consult with care coordination staff and participants' health care providers, and provide services and referrals.

  This leadership change also resulted in a change in the structure of mental health promotion services for the remainder of the award period. Under this new structure, the range of services included preventive offerings (for all participants) such as one-on-one or online education in stress management and self-care, consultation (for one-time requests from community health workers or participants' health care providers), early intervention (for

participants who need support like grief counseling but would likely not meet the diagnostic criteria at a community mental health center), and referrals (for participants in need of longer-term mental health treatment or services closer to home).

Program staff reported that mental health services expanded considerably in the third year of the program, as the new staffing model along with the implementation of health technology (for example, the videoconferencing platform) facilitated work with participants. Staff conducted group psychoeducation sessions at different locations in the community, met with participants at the program offices, and used secure two-way text messaging and videoconferencing to communicate with participants remotely.

- **Health technology.** The program had a team of four IT specialists working with a number of contracted technology vendors on care coordination software and other health technology to support the CHECK program. These program staff worked with a technology vendor to customize an existing care coordination software product to the needs of the CHECK program. The care coordination software allowed community health workers to document assessments, generate care plans, and track ongoing contacts with participants and their families. The software also included an integrated repository of social service resources that community health workers used to provide and track referrals. In addition, CHECK offered health technology tools such as online self-education portal generated by the program and a two-way text messaging platform, as well as videoconferencing technology to support participant engagement in the program. The videoconferencing and two-way text messaging platforms were developed and introduced in the second year of program, after program leaders and staff identified the need for more flexible means for participants to communicate with care coordination and mental health staff.

> "At the beginning, we needed to be really focused on assessments. And providing great quality care can't just be about the assessments. Last August [2016], we hit the mark that we were supposed to. So this year, we've been able to really focus on engaging the patients that we have, making sure that we've engaged those patients and we're continuing to engage the patients and work with them, making sure we're providing quality care, helping them with appointments, and things like that."
>
> —Program leader

b. Evidence of service delivery effectiveness

Overall, we assess the University of Illinois to be partly successful in implementing the service delivery model. Our assessment is based on delays in implementing the program's core care coordination and mental health promotion services but eventual success implementing them consistently by the third year of the program.

Delivery of intervention services. In site visit interviews, program leaders and staff reported that CHECK program services were somewhat slow to ramp up over the course of the award period, but by the end of the award period, the program was consistently delivering its core intervention services. Leaders and staff noted that the majority of the first year of the program was dedicated to developing plans for program services, building technology tools, and working through bureaucratic challenges posed by being part of a large state university system. During the second year of the program, leaders and staff noted that the focus shifted from

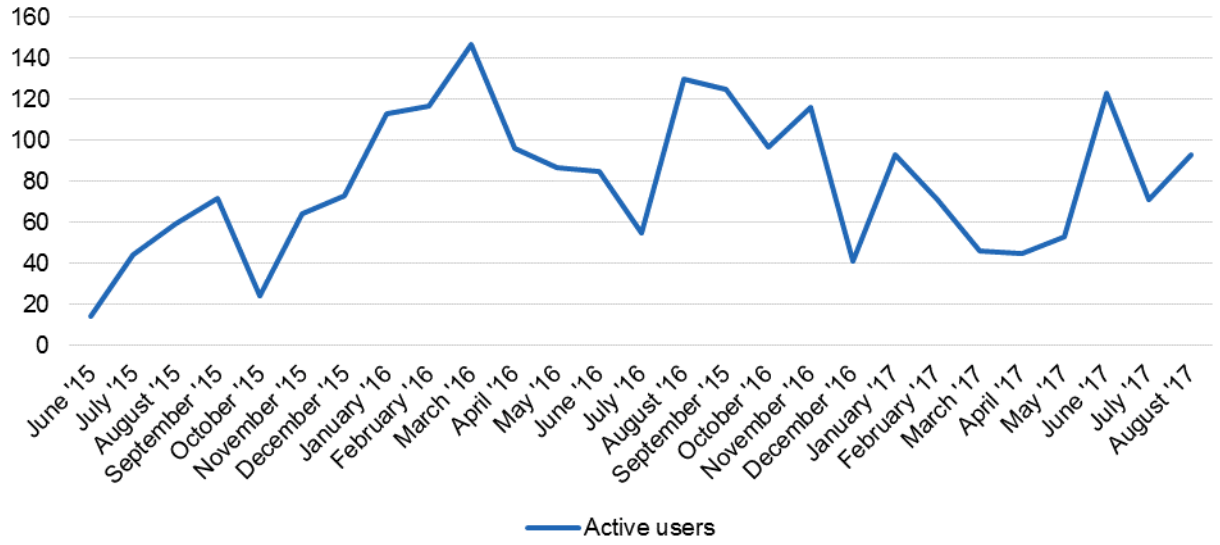
"starting up" to engaging participants in the program. Staff also continued to face some challenges balancing initial outreach and assessments with ongoing services to participants.

Additionally, in the second year, the CHECK program also underwent a significant change in leadership within the care coordination and mental health promotion team. The new leaders restructured the program's service delivery teams, staff oversight, and training, and redesigned their protocols. By the third year of the program, CHECK staff reported they were successfully delivering ongoing care coordination and mental health services. This evolution in program operations was corroborated by responses to the staff survey fielded in July to October of 2016. Nearly two-thirds of staff who responded to the survey (65 percent) agreed that the work environment had been chaotic at the beginning of the program, but less than one-third of respondents (28 percent) thought the environment was still chaotic at the time of the survey, around the start of the third program year. In addition, 88 percent of staff who responded to the survey noted that program logistics improved over time.

The awardee's self-reported monitoring data provide more evidence of the CHECK program's progress in delivering services. According to these data, about 97 percent of the participants who were ever engaged in the program had a care plan created in the care coordination software system—exceeding the program's goal of 90 percent. This suggests that once staff were able to find and communicate with participants, these participants got an assessment and had their care needs identified and documented.

The number of participants who received social service referrals (for example, to food programs, mental health providers, or immigration or housing advocacy) steadily increased during the first few months, with some seasonal spikes in summer and winter over the following months (Figure II.1).The number of participants who actively requested social service referrals ranged from one to 147 in the period during which the awardee reported data on this metric. Requests for referrals grew steadily until March 2016; after that, they fluctuated. On average, the program provided two referrals per participant (based on those who requested referral services) each month. Over the course of the reporting period, the referral rate ranged from low of .1 referral per active user in the first month of reporting to a high of 3.7 referrals per active user in February and March 2017.

Figure II.1. Social service referrals to CHECK



Note: Active users are defined as the unique number of CHECK participants who received social service referrals each month. Referrals were provided and tracked through the program's social service repository, integrated within the care coordination software.

This suggests that community health workers made use of program technology to meet participants' needs.

The awardee did not report other metrics on the ongoing services it delivered, limiting our ability to assess effectiveness in delivering services beyond the initial assessment and the number of electronic social service referrals.

Staffing and training. The University of Illinois had some trouble hiring and retaining staff for CHECK in the first two years, but this issue was resolved by the beginning of the third year. In the first year of the cooperative agreement, hiring was delayed because the CHECK program is housed in a public university, and the awardee must adhere to the state's hiring and procurement requirements. Program leaders estimated that these requirements added three to six months to the hiring process, impeding implementation in the early stages. Specifically, the awardee had to delay hiring the care coordination staff, who were essential to engage and serve program participants.

Data reported by the awardee suggest that even as the number of full-time staff increased, staff retention was an issue in the first two years of the program, with some turnover reported in nearly all quarters. Survey data support this finding, with almost two-thirds of staff survey respondents in fall 2016 believing that staff turnover was a major (19 percent) or minor (47 percent) barrier to program effectiveness.

The awardee reported being fully staffed as of August 2016, two years after the program began, with 71 staff in place, and also reported minimal turnover in the third year of the program. At the end of the three-year cooperative agreement, program leaders reported that they chose to scale back staffing to improve sustainability as their fourth year with a no-cost extension began. The program appears to have a significantly smaller staff in its final quarter, going from 70 staff

in May 2017 to 45 staff in August 2017. The reduction included all types of staff, but focused on management and community health workers.

The awardee successfully implemented its training plans for CHECK staff. In site visit interviews in the third year of the program, program leaders reported that all community health workers had received training on engaging patients and coordinating care. The CHECK program developed care coordination protocols for managing the four targeted conditions (asthma, diabetes, prematurity, and sickle cell disease) as well as for oral disease and mental health promotion. Each community health worker received at least one disease-specific training, and all community health workers received training in asthma treatment because this condition was so prevalent among CHECK participants and family members.

Given the small size of the mental health promotion team, the awardee provided most training for these staff through less formal means, such as frequent supervision to discuss program protocols and specific participant cases. In addition to role-specific trainings, the CHECK program provides broader trainings for program staff on topics such as cultural competency (for example, working with immigrant populations). During site visit interviews, both care coordination and mental health promotion staff described their training and/or supervision as useful and supportive for their role in the program.

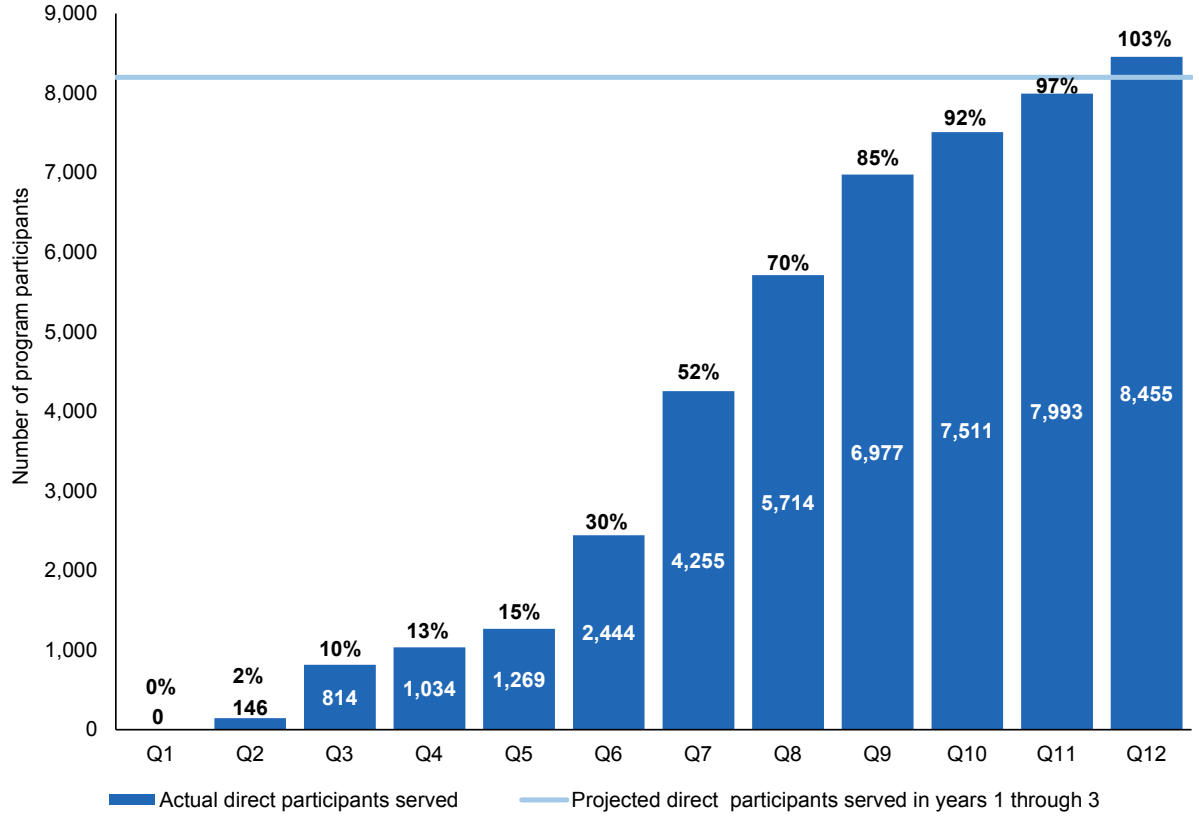
Almost all staff members who responded to the survey in fall 2016 (97 percent) said they attended at least one formal training during the award period. About 9 in 10 staff respondents said the trainings gave t hem new skills important for their role (89 percent), and almost as many (85 percent) reported that the trainings improved their job performance within the program. However, more than half of the respondents (53 percent) wished they had been offered other trainings related to their role in the program. Most staff also reported participating in various types of informal staff training, such as through staff meetings, team huddles, and individual and group supervision.

Recruitment and engagement of providers. The awardee did not recruit providers to implement the model. Therefore, we did not assess implementation effectiveness on the basis of provider recruitment and engagement.

Engagement of program participants. After enrolling participants by sending them a letter telling them about the CHECK program, community health workers used telephone and in-person outreach to engage participants. The awardee considered participants to be engaged in the program once a community health worker had made at least one contact with the participant or family member—most often to initiate a care plan.[7] There were some initial struggles to engage participants because of inaccurate contact information, difficulty connecting with the participant population, and challenges managing staff workload; these issues are described in detail below. However, the awardee ultimately exceeded its goal for engaging participants, reporting that the program served 8,455 direct participants by the end of the initial three-year award period (Figure II.2).

[7] In addition to "enrollment" and "engagement," the awardee defined an additional level of program participation—"activation"—which referred to participants for whom community health workers completed a care plan in the care coordination software. Staff reported that they did not make this distinction in practice; as a result, we do not discuss activation separately from engagement when assessing implementation effectiveness.

Figure II.2. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Participants served data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. Participants are considered to have been served by the program if they received at least one contact from CHECK staff, including an initial assessment.

Program leaders and staff noted that CHECK's ability to engage participants improved over the course of the program. By the time of the staff survey in fall 2016, three-quarters (75 percent) of staff who responded to the survey agreed that the awardee succeeded in engaging program participants. In addition, participant engagement was the program facilitator cited most often, with nearly half of staff (46 percent) saying that it positively impacted the implementation's effectiveness.

In site visit interviews, many staff attributed the enhanced levels of engagement to the changes in the care coordination team structure during the second year of the program. For example, the CHECK program created a separate, dedicated team of community health workers focused solely on going out into the community to find participants who were difficult to reach or who had fallen out of touch with the program. As one program leader noted, "The community health workers know their communities really well, so they're in a really good position to know where people might be … living or whatever is going on."

Staff noted, however, that even after making initial contact with the program, some youth and families declined to participate. Survey and interview findings suggest that participants may have found the program to be confusing or burdensome. Nearly two-thirds of the staff who responded to the survey (64 percent) believed that participants who opted out of the program did so because they did not understand it. More than half (56 percent) reported that participants opted out because there was too much required from them. In site visit interviews, staff said the participants sometimes confused the program with other care coordination initiatives, such as those offered directly by federally qualified health centers or MCOs, or questioned why their family would need or benefit from the program.

c. Barriers and facilitators associated with service delivery effectiveness

The University of Illinois faced several ongoing barriers to implementing the CHECK program, including (1) the broader university's hiring and contracting policies, (2) technology that did not meet service delivery needs, (3) management of staff workload, and (4) a hard-to-reach target population.

Hiring and contracting policies. Program leaders reported that the university's institutional policies and requirements impeded implementation, noting that these policies significantly delayed their hiring of program staff and executing of contracts with technology vendors and managed care plans during the first year of the program, and that these delays persisted throughout the award period.

Program technology. At every annual site visit, program leaders and frontline staff said the program's care coordination software was one of the biggest barriers to implementation. During the first year of the program, as CHECK staff began enrolling participants identified through claims data, staff reported that the care coordination software was not functioning as it was supposed to, requiring care coordination teams to develop their own ad hoc documentation strategies. This placed a significant burden on staff as they adjusted to their new positions and worked to implement program protocols for engaging participants and following up with services.

> "The really big thing that we've learned in the last year is how much these families don't know about the resources that are available to them [in their communities]. And how the CHWs are really like an information repository of resources. And especially because they live in the areas where the patients do, they know on a zip-code basis what's available there."
>
> —Program leader

Although program leaders worked closely with their technology vendor to improve the software over the course of the program, staffing changes at the vendor drew out the process. In site visit interviews, program leaders noted that the software ultimately never functioned the way it was supposed to, even after it was improved. At the time of the last site visit, program leaders were pursuing other options for care coordination software in partnership with the university health system.

Staff workload management. The program staff sometimes struggled to balance outreach to newly enrolled (but not yet engaged) participants with the ongoing provision of services to participants and their families. The same staff members (community health workers) were tasked with both activities, and during the first year as they worked to build a participant base, staff said they were overwhelmed by competing demands. To resolve this challenge, in the second year of

the program, the awardee used a call center affiliated with its care coordination software provider to make the initial phone calls to participants who were enrolled in the program but not yet engaged. This allowed the awardee to keep building its participant base while giving community health workers more time to focus on doing outreach in the field, building relationships with participants and families, and giving engaged participants the services in their care plans.

Program leaders and frontline staff reported that the care coordination teams charged with enrolling and engaging participants became adept at balancing these competing priorities over the course of the cooperative agreement, allowing staff to invest more time and effort in engaging participants. In interviews, program leaders described using multiple data- and staff-driven approaches to foster this, such as sharing progress toward engagement goals through posters in work areas and encouraging staff members to leverage their different strengths and comfort levels to maximize productivity (for example, if they had a preference for phone versus face-to-face engagement with participants). Program leaders reported that, as of the end of the third year, program staff were working at capacity, dividing their time between engaging and activating enrolled participants and providing ongoing services.

Hard-to-reach population. Despite the ultimate success in meeting their goals for engaging participants, care coordination staff said it could be hard to find and engage participants throughout the three-year cooperative agreement period. In site visits, many frontline staff noted that participants changed their phone numbers and addresses often, or that the managed care data the program received often had outdated contact information As a result, the program staff devoted significantly time and effort than expected to making initial contact with participants and their families.

Moreover, participants' families may have been wary of the program's outreach, particularly if the family were concerned about government intervention in their lives (for example, a family may have feared that CHECK staff were from an immigration or child welfare agency). Although program staff reported that inaccurate contact information was a challenge for the whole length of the program, they also noted that the program's approach to prioritizing and managing this outreach improved over time. For example, program leaders developed and refined protocols for staff to prioritize their contact efforts and decide when to stop doing outreach. CHECK program leaders also implemented several adaptations to standard practices to meet observed participant needs, including creating a dedicated care coordination team focused solely on hard-to-reach participants, and shifting some care coordination staff to evening or weekend shifts when they might be more likely to reach participants.

The awardee also identified several strategies and strengths that helped the program to overcome these barriers, including (1) building relationships with external partners, (2) identifying and deploying effective leaders, and (3) creating a strong workplace culture.

Relationships in the health system and community. CHECK program leaders noted that the program's increasingly close relationships within the university hospital system and with outside community health centers helped get more participants enrolled and kept them engaged. From the program's outset, leaders worked to establish relationships with community- and school-based health centers. Late in the second year of the program, program leaders began assigning community health workers to some of these health centers in order to (1) connect with

CHECK participants in a convenient location before or after their medical appointments, and (2) collaborate with clinic staff to promote the program and encourage new referrals.

CHECK also has program leaders who focus on involving the community, and work with liaisons, or "champions," at each site to identify the sites' needs and the ways that CHECK's community health workers could be most useful. CHECK program leaders also built relationships in the University of Illinois hospital and identified ways for the program to engage participants when they were admitted. For example, dedicated CHECK community health workers attend clinical rounds in each hospital unit, and CHECK's community engagement leaders regularly attend meetings with hospital leaders from these units. In interviews, care coordination staff and leaders said having access to participants in this setting was extremely valuable, especially in the process of supporting discharge planning. These activities helped identify new participants for the program and engage existing participants more effectively. They also laid the foundation for future partnerships with the hospital and community health centers.

> "We put a lot of time and energy into making this a fun place to work, a place where you can experience positivity with your colleagues. The things that they're dealing with are so intense, so heavy. Just listening to some of these stories and the level of trauma and turmoil that these families have been through is really challenging. I think a lot of [staff] feel like they are in a good working environment."
>
> —Program leader

Operational leadership. CHECK program administrators and frontline staff pointed to the change in leadership of the care coordination and mental health promotion teams early in the second year of the program as a turning point. These program leaders developed and deployed new program protocols, clarified roles, and provided more training, which in turn improved staff's ability to engage participants in services. As part of these changes, the program leaders initiated changes in the structure of care coordination and mental health promotion teams to better meet identified needs. The new model allowed for greater collaboration within each team and adaptability to the target area each team was assigned to. The simultaneous redesign of the mental health team, allowing for a more centralized team of full-time mental health staff, helped the program to consistently conduct a broader range of services, including assessments, consultations with the care coordination teams, brief intervention, and referrals.

Workplace culture. In the second and third years of the program, program leaders said they put more emphasis on making the program a collaborative working environment, consistently gathering input from frontline staff. This collaborative environment promoted the use of continuous quality improvement strategies, which helped them adjust protocols and policies. To this end, several program leaders were trained in Lean Six Sigma performance improvement methodology, and they incorporated the lessons they learned about feedback and transparency into CHECK. For example, care coordination staff noted that program leaders invited them to make suggestions and comments on a bulletin board in a common area in the CHECK offices. In site visit interviews, both program leaders and frontline staff reported that this input helped make the program more responsive to the needs of staff and participants.

The awardee's increased emphasis on collaborative processes and workplace culture may have contributed to a strong sense of satisfaction on the part of staff. Three-quarters of staff responding to the survey said they felt supported by colleagues and management, and the same number reported being somewhat or very satisfied with their current role in the CHECK

program. Staff also cited support from program leaders (37 percent); effective communication (31 percent); and sufficient staffing, financial resources, and staff buy-in (all 26 percent) as being factors in the program's success.

Survey findings suggest, however, that some of the workload issues that affected frontline staff in the program's first year may persist. In the survey conducted in the fall of 2016, shortly after implementation of the staffing redesign, 53 percent of staff reported that their workload was somewhat or very heavy, and 50 percent reported that their position with CHECK increases their feelings of workplace burnout. About half of the staff respondents (48 percent) believed that not having enough time to do their work was a barrier to program effectiveness.

C. Assessment of perceived program effects on the delivery of care and outcomes

The majority of staff who responded to the staff survey thought the CHECK program had positive impacts on the delivery of care and outcomes for participants and their families (Table II.2). Eighty-one percent of respondents to the staff survey reported that CHECK was "somewhat effective" or "very effective" in achieving program goals. When asked to indicate what had made the program effective in reaching its goals, staff cited better treatment plans for participants, leading to improved outcomes (33 percent), improved participant awareness and education (20 percent), and improved participant engagement (20 percent).

Staff indicated that the program involved intensive outreach efforts to reach potential participants and keep current participants engaged in the program, including home visits (64 percent), telephone outreach (67 percent), and community event-based outreach (70 percent) in addition to recruitment and enrollment efforts (67 percent).

Table II.2. Staff perceptions of CHECK's effects on care

Percent of non-clinician staff indicating that the CHECK program had a positive impact on the following:	Percentage of applicable respondents (n = 36)[a]
The quality of care and services you provide to participants[a]	97
Your ability to respond in a timely way to participant needs[a]	91
The efficiency of care services provided to participants[a]	94
Your ability to provide care or services that are responsive to participant preferences, needs, and values[a,b]	97
Care services that are provided fairly to all participants[a,b]	91
Access to care or services for all participants	94
Achievement of participants' health goals [a]	91
Participant satisfaction[a]	94
Participant quality of life[a]	94
Care coordination[a]	94

Source: HCIA R2 evaluation survey of participating non-clinician staff, fall 2016.

[a]Several respondents in the 36 returned surveys indicated that these items were not applicable to their position. These were not included in the calculations here.

[b]One missing response.

In interviews, different staff said they believed in the positive effects the program had on participants' lives, citing its focus on social determinants of health to drive what was described in interviews as a "trickle-down" of program benefits: that is, the effects of the intervention will affect areas of the participants' lives not directly addressed by the program's work. "It's not about giving them [the resources]," one frontline staff member said. "We're teaching them something that they'll take with them."

Seventy percent of staff in the survey said they help program participants access non-medical services such as housing, job training, or legal assistance; and 75 percent report providing education to participants about managing their health care and other services. In interviews, staff members describe using the immediate needs of the participants, such as legal assistance or connections to mental health services, to earn the families' trust and build an engaging relationship before addressing the participants' medical complexity.

Program leaders and staff said the education participating families got established a baseline knowledge to help them navigate the health system in the future as adults with chronic medical conditions. The program, they believe, addresses the core issues affecting the health and well-being of children with medical complexity, such as housing stability, legal assistance, and family mental health.

At the same time, program leaders and staff have reservations about the program's ability to produce measurable impacts in the short term. Because the program design focuses on social determinants of health, they believe the improvements to participants' lives will take longer to appear. They note the complexity of outcome measures like school attendance and family mental health and quality of life, explaining that these cases require a number of diverse interventions from the CHECK program. Improving performance on these measures, from their perspective, will require long-term, gradually mounting participant engagement. Frontline staff did note that the length of the three-year cooperative agreement allowed them to see initial results from early program enrollees, validating their belief in the importance of building sustainable relationships with participants.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. After initial challenges, the program seemed to be operating more smoothly by the end of its second year (according to feedback on site visits and surveys). By this time, new leaders of the care coordination and mental health promotion teams had implemented changes that led to better service delivery, staff oversight and training, and protocols. If CHECK was indeed operating more consistently by the time of the staff survey (July to October 2016, around the end of the second year of the program), then the time frame for our impact evaluation will include a period of partial implementation in the second year and full implementation in the third year. We would expect any impacts to reflect those stages of implementation.

As noted, findings from site visit interviews with program leaders and staff about the effects of the intervention on the health and well-being of participants and their families suggest that over time the program has the potential to reduce health care costs and improve health outcomes.

Staff described the challenges of working with a target population with diverse and complex medical and non-medical needs, recognizing that although immediate needs could be met by the program (such as help with paperwork or referrals to community services) most other needs (such as mental health issues) would take time to address. These staff did not expect the program's long-term positive impacts to be measurable within the time frame of our evaluation.

Our evaluation focuses on CMMI's core measures for outpatient ED use, hospital admissions, and total Medicaid spending. The program is also intended to improve other key outcomes, such as engagement in services, quality of life, and school attendance, but data to measure these outcomes are not available. In the following sections, we report impacts on the CMMI core measures over the 12-month period from April 2016 to March 2017.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents our preliminary findings for the RCT in the first year after randomization. We present baseline characteristics of the treatment group, an assessment of the balance between treatment and control groups on baseline characteristics, and impact estimates for the first year of the RCT.

Key findings:

- CHECK reduced the number of outpatient ED visits in the first year of the RCT among the full population. Most of the reduction occurred within the first six months after randomization.
- CHECK reduced the number of hospitalizations and inpatient and total spending in months 7–12 after randomization in the medium- and high-risk subgroup.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify the treatment and control groups by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of September 1, 2017: University of Illinois

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018, in randomized sample	3,131[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,649
Likelihood of all-cause hospitalizations	1,355
MDE sample size requirement to detect 20% effect	
Total expenditures	412
Likelihood of all-cause hospitalizations	339
Participation/selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable control group?	Yes, high rate of identification of treatment and control groups with claims data within an intent-to-treat framework
Likelihood of solid control group	No serious issues; proceeding with control group
Do claims identify the primary expected effects?	Yes

Table III.1 *(continued)*

Evaluability domain	Response
Core outcomes estimation method	Regression-adjusted comparison of treatment and control group
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	Possible data on use of care coordination services

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We are conducting a rigorous impact analysis by comparing outcomes for the randomly assigned treatment group to outcomes for the randomly assigned control group, for both the full randomly assigned population as well as a subgroup of medium and high-risk beneficiaries. The projected sample size in Table III.1 reflects the total number of treatment group beneficiaries in the randomized sample. Preliminary findings from our impact evaluation of the program effect on a selected set of outcomes are included in this report. We have a sufficient sample of randomly assigned beneficiaries to be able to detect plausible effects on claims-based measures.

B. Characteristics of RCT treatment group at baseline

A total of 6,259 Medicaid-eligible children in Harmony Health Insurance Plan (Harmony), an Illinois managed care organization, were included in the randomized sample.[8] Mathematica randomly assigned 3,128 of these beneficiaries to the control group and 3,131 to the treatment group by using a randomization procedure that was stratified by the beneficiary's risk tier (high, medium, or low), target condition (asthma, diabetes, prematurity, or sickle cell disease), age group (birth to age 5, 6 to 10, 11 to 15, or 16 to 25) and zip code. Randomization took place on April 11, 2016. The baseline period for the analyses covers the 12-month period from April 1, 2015, to March 31, 2016. We used full calendar months to define the baseline and intervention periods to correspond with monthly Medicaid eligibility data.[9]

Table III.2 shows the baseline characteristics of the treatment group, both for the overall RCT population and the subgroup of medium- and high-risk participants. Although the average age at the start of the baseline period was just over 11 in both groups, the distribution of age differed. The medium- and high-risk subgroup had more beneficiaries from ages 19 to 25 than

[8] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

[9] We later learned that no Medicaid eligibility data for the treatment group were available for the baseline period (based on communication with Joanna Tess, monitoring and evaluation manager at CHECK, August 30, 2017). When measuring baseline characteristics from Medicaid claims data, we assumed that all treatment and control beneficiaries were enrolled for the full 12-month baseline period.

the full population did (24 percent versus 15 percent) and fewer beneficiaries ages 9 to 11 (11 percent versus 16 percent) or 12 to 18 (22 percent versus 31 percent). Fifty-three percent of the full group were male, and 50 percent were male in the medium- and high-risk subgroup.

Asthma was by far the most common target condition among participants, with 86 percent of the full group and 84 percent in the medium- and high-risk subgroup having this diagnosis. Nationally, 8.4 percent of children under age 18 and 7.6 percent of adults age 18 and older have asthma (Centers for Disease Control and Prevention 2017a). Six percent had diabetes in both the full group and the subgroup. Those rates are over 20 times the 2015 national diabetes rate of 0.24 percent found in people under age 20 (Centers for Disease Control and Prevention 2017b). One percent had sickle cell disease in both the full group and the subgroup. In the United States, sickle cell disease is most prevalent among African Americans. Nearly one in 365 black children born in the United States has sickle cell disease (National Heart, Lung, and Blood Institute 2017). The prematurity rate was 5 percent in the full group and 6 percent in the medium- and high-risk subgroup.

Looking at the CHECK program's secondary chronic conditions, a diagnosis of newborn brain injury was evident among 0.4 percent of the full treatment group and 1 percent of the subgroup of medium- and high-risk beneficiaries. The percentage of CHECK participants with a seizure or epilepsy was 4 percent in the full group and 5 percent in the medium- and high-risk subgroup. Nationally, about 1 percent of Americans have epilepsy across all age groups (Centers for Disease Control and Prevention 2017c).

In the full treatment group, 4 percent were flagged as high risk, 22 percent as medium risk, and 74 percent as low risk. By definition, the group of participants called the subgroup here was composed only of the medium- and high-risk beneficiaries. We also constructed a Chronic Illness and Disability Payment System (CDPS) score to proxy for health status.[10] The average CDPS score among the full treatment group was 2.7, suggesting that expected Medicaid spending was 2.7 times higher than average spending for the population of children on Medicaid whose families had low incomes. A higher-than-average score for the treatment group was expected, because all participating children had at least one chronic condition. The average CDPS score was 3.6 in the medium- and high-risk subgroup, suggesting that expected spending was even higher for the subgroup than for the full population of participants.

About 5 percent of the full treatment group had any hospitalizations during the baseline period, and 34 percent had any outpatient ED visits. The rates were higher in the subgroup, where 17 percent had any hospitalization, and 73 percent had any outpatient ED visit. Average spending per beneficiary per month in the full group was $183. Of this, inpatient spending accounted for 39 percent, outpatient ED visits accounted for 5 percent, prescription drug

[10] The CDPS was developed by researchers at the University of California, San Diego to adjust Medicaid payments to managed care organizations. The CDPS score is based on demographic characteristics, the presence of specific diseases and conditions that are characterized by expected cost (ranging from extra high to very low), and types of prescription drug scripts. A higher CDPS score signifies a higher expected risk profile and higher expected spending, whereas a lower CDPS score signifies a lower expected risk profile and lower expected spending. For all beneficiaries in the RCT, we calculated scores using a "TANF child population" model (Kronick et al. 2000). The scores are normalized so that the average spending in each population is one. A score above one indicates higher than average expected spending, and a score below one indicates lower than average spending.

spending accounted for 16 percent, and all other Medicaid-covered services accounted for the remaining 40 percent. Average PBPM spending in the subgroup, $372, was about twice as high as in the full group. Of that spending, inpatient spending accounted for 51 percent, outpatient ED visits for 7 percent, prescription drug spending for 10 percent, and all other services for the remaining 31 percent.

We added zip-code–level characteristics to the analysis file using publicly available data from the American Community Survey. The distribution of zip-code–level characteristics was similar for the full group and the subgroup. For example, the average percentage of white residents was 26 percent for the full sample and 30 percent for the subgroup. The average percentage of black participants was also comparable: 38 percent for the full sample and 35 percent for the subgroup, and the average percentage of Hispanics was 32 percent for the full sample and 30 percent for the subgroup.

Table III.2. Baseline characteristics of Medicaid FFS and managed care beneficiaries in the randomized treatment group: full group and subgroup of medium- and high-risk beneficiaries, University of Illinois

Measure	Treatment group: all beneficiaries (n = 3,131)	Treatment group: medium- and high-risk subgroup (n = 821)
Age and gender		
Age at the end of the baseline year (continuous)	11.2	11.5
Age group, %		
Birth to age 8	39	43
9–11	16	11
12–18	31	22
19–25	15	24
Male, %	53	50
Target conditions, %		
Asthma	86	84
Diabetes	6	6
Prematurity	5	6
Sickle cell disease	1	1
Secondary conditions, %		
Newborn brain jury	0.4	1
Epilepsy/seizure	4	5
Risk tier, %		
High	4	17
Medium	22	83
Low	74	N/A

Table III.2 *(continued)*

Measure	Treatment group: all beneficiaries (n = 3,131)	Treatment group: medium- and high-risk subgroup (n = 821)
Health status and Medicaid utilization and expenditures pre-RCT		
CDPS score	2.7	3.6
Any hospitalizations, %	5	17
Number of hospitalizations	0.07	0.23
Any outpatient ED visits, %	34	73
Number of outpatient ED visits	0.62	1.6
Any asthma hospitalizations, %	1	2
Number of hospitalizations for asthma	0.01	0.02
Any outpatient ED visits for asthma, %	5	12
Number of outpatient ED visits for asthma	0.06	0.18
Total Medicaid expenditures ($ PBPM)	$183	$372
Inpatient expenditures ($ PBPM)	$71	$191
Outpatient ED expenditures ($ PBPM)	$9	$26
All other medical expenditures ($ PBPM)	$73	$117
Prescription drug expenditures ($ PBPM)	$30	$39
Zip-code–level characteristics		
White, %	26	30
Black, %	38	35
Hispanic, %	32	30
Other race/ethnicity, %	5	5
Median income	$45,810	$48,069
Poverty, %	23	22
High school degree or higher, %	77	78
College degree or higher, %	20	21
Unemployed, %	15	14
U.S. citizens, %	89	89
Uninsured, %	17	17

Notes: The medium- and high-risk subgroup comprises patients who had one to three emergency department (ED) visits or one hospitalization in the year before randomization (medium-risk) or more than three ED visits or more than one hospitalization in the year before randomization (high-risk).

C. Program participation among the treatment group

The University of Illinois engaged or activated 23 percent of the treatment group beneficiaries in CHECK within the first quarter after randomization (April–June 2016). The share of engaged or activated beneficiaries grew to 25 percent by the fourth quarter of the first year of the RCT (January–March 2017). Most patients who were activated or engaged in the first quarter stayed engaged or activated throughout later quarters. The pattern of engagement and activation was similar for the medium- and high-risk subgroup. The relatively small share of the treatment group that was engaged or activated may limit our ability to detect impacts.

D. Quality of randomization

The results in this section demonstrate that randomization worked well. Covariates are well balanced between the treatment and control groups, both in the full population and in the subgroup of medium- and high-risk beneficiaries.

1. Checking balance

a. Variables

We included a number of beneficiary characteristics in the balance assessment: demographic characteristics, target conditions, secondary conditions, risk tier, and measures of health status and utilization and spending (Table III.3).

Table III.3. Variables used for checking balance, University of Illinois

Variable	Data source
Demographics	
Age at the end of the baseline year	UIC randomization file
Gender	UIC randomization file
Target conditions, secondary conditions, and risk tier[a]	
Asthma	UIC randomization file
Diabetes	UIC randomization file
Prematurity	UIC randomization file
Sickle cell disease	UIC randomization file
Newborn brain injury	UIC randomization file
Epilepsy/seizure	UIC randomization file
Risk tier[a]	UIC randomization file
Health status, utilization, and spending in the baseline year	
CDPS score[b]	Claims data
Any hospitalizations	Claims data
Number of hospitalizations	Claims data
Any outpatient ED visit	Claims data
Number of outpatient ED visits	Claims data
Any asthma hospitalization	Claims data
Number of asthma hospitalizations	Claims data
Any asthma outpatient ED visit	Claims data
Number of asthma outpatient ED visits	Claims data
Total Medicaid expenditures ($ PBPM)	Claims data
Inpatient expenditures ($ PBPM)	Claims data
Outpatient ED expenditures ($ PBPM)	Claims data
All other medical expenditures ($ PBPM)	Claims data
Prescription drug expenditures ($ PBPM)	Claims data

Table III.3 *(continued)*

Variable	Data source
Zip-code–level characteristics	
Percent white, black, Hispanic, or other	American Community Survey
Median income	American Community Survey
Poverty rate	American Community Survey
Educational attainment	American Community Survey
Unemployment rate	American Community Survey
Percent U.S. citizens	American Community Survey
Percent uninsured	American Community Survey

[a]Low-risk participants had no emergency department (ED) visits and no hospitalizations in the year before randomization. Medium-risk participants had one to three ED visits or one hospitalization in the year before randomization. High-risk participants had more than three ED visits or more than one hospitalization in the year before randomization.

[b]We calculated CDPS scores by using the software from version 6 to flag condition-cost categories and prescription drug categories. We used the weights from version 5.4 of the software. There was no change in the weights between versions 5.4 and 6. The difference between the versions was that version 5.4 only uses ICD-9, whereas version 6 allows for both ICD-9 and ICD-10 codes, and it does not create some condition-cost categories that were excluded from the models for calculating the final score. For all participants, we used the CDPS+MRX algorithm for TANF children. The CDPS was developed by researchers at the University of California, San Diego.

CDPS = Chronic Illness and Disability Payment System.

2. Balance results

Figure III.1 shows the balance between the full treatment group and the full control group in terms of standardized differences of important baseline variables, with boundaries at 0.10 and 0.25 above and below zero. No variables had an absolute standardized difference greater than 0.10, suggesting that the treatment and control groups are well balanced. Appendix Table A.1 provides more data on the balance between the two groups. The omnibus test for balance (p = 0.89) gives overall assurance that the groups do not differ significantly on the measured characteristics. The p-values for each row show that there are no significant differences between groups on any measured characteristics.

The subgroup of medium- and high-risk patients also appears well balanced between the treatment and control groups. Figure III.2 shows that no variables had an absolute standardized difference greater than 0.10. Appendix Table A.2 shows that only one variable was significantly different between the treatment and control beneficiaries. Specifically, the share of beneficiaries with a diagnosis of newborn brain injury was 1 percent in the treatment group compared to 0.1 percent in the control group (p = 0.034). However, in both groups, the prevalence of this condition is low, and this difference does not change our conclusion that the two subgroups are generally well balanced. The omnibus test for balance (p = 0.65) provides further assurance that the two groups do not differ on the measured characteristics.

Figure III.1. Standardized differences between treatment and control beneficiaries on baseline variables: full sample, University of Illinois

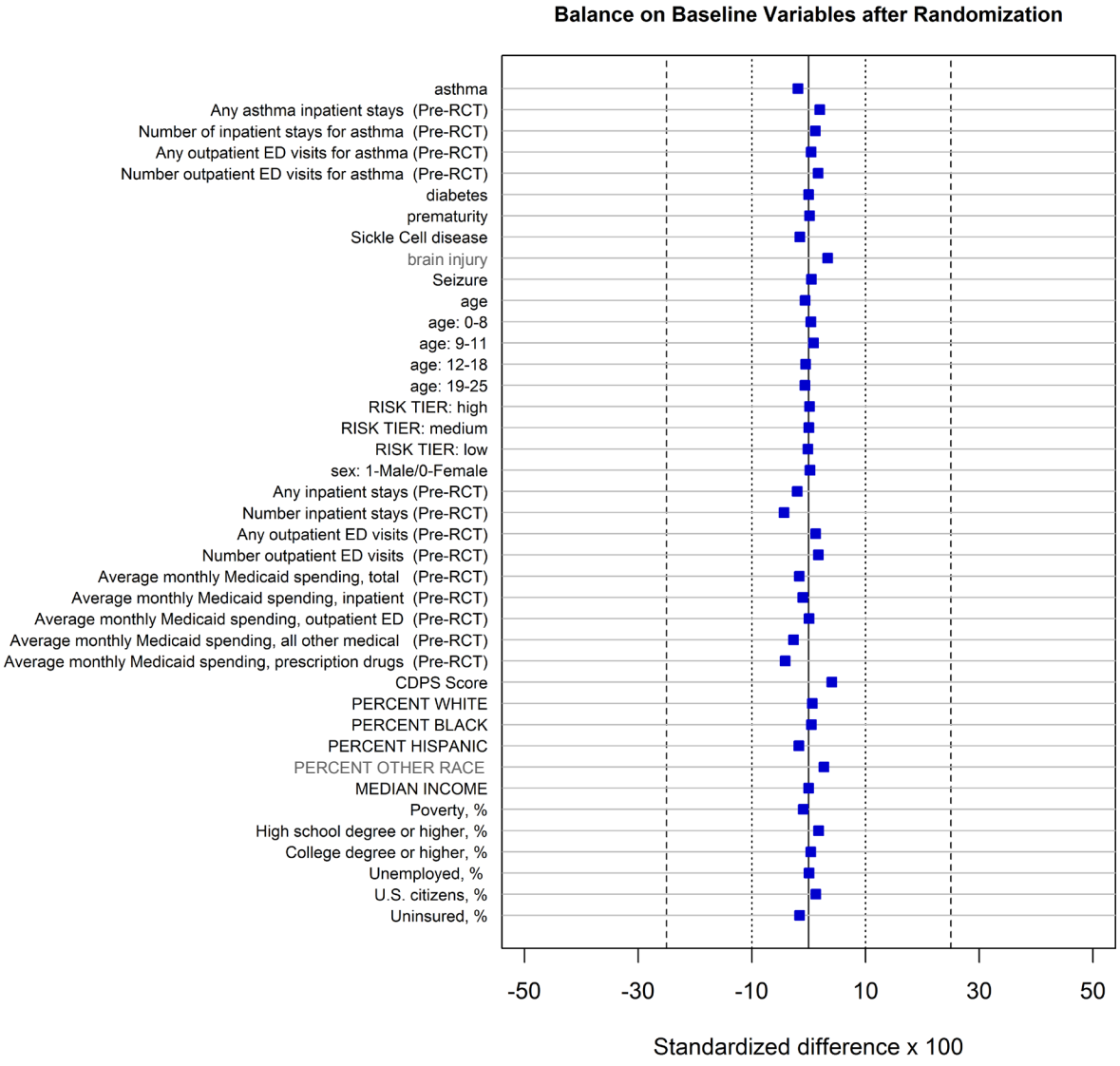
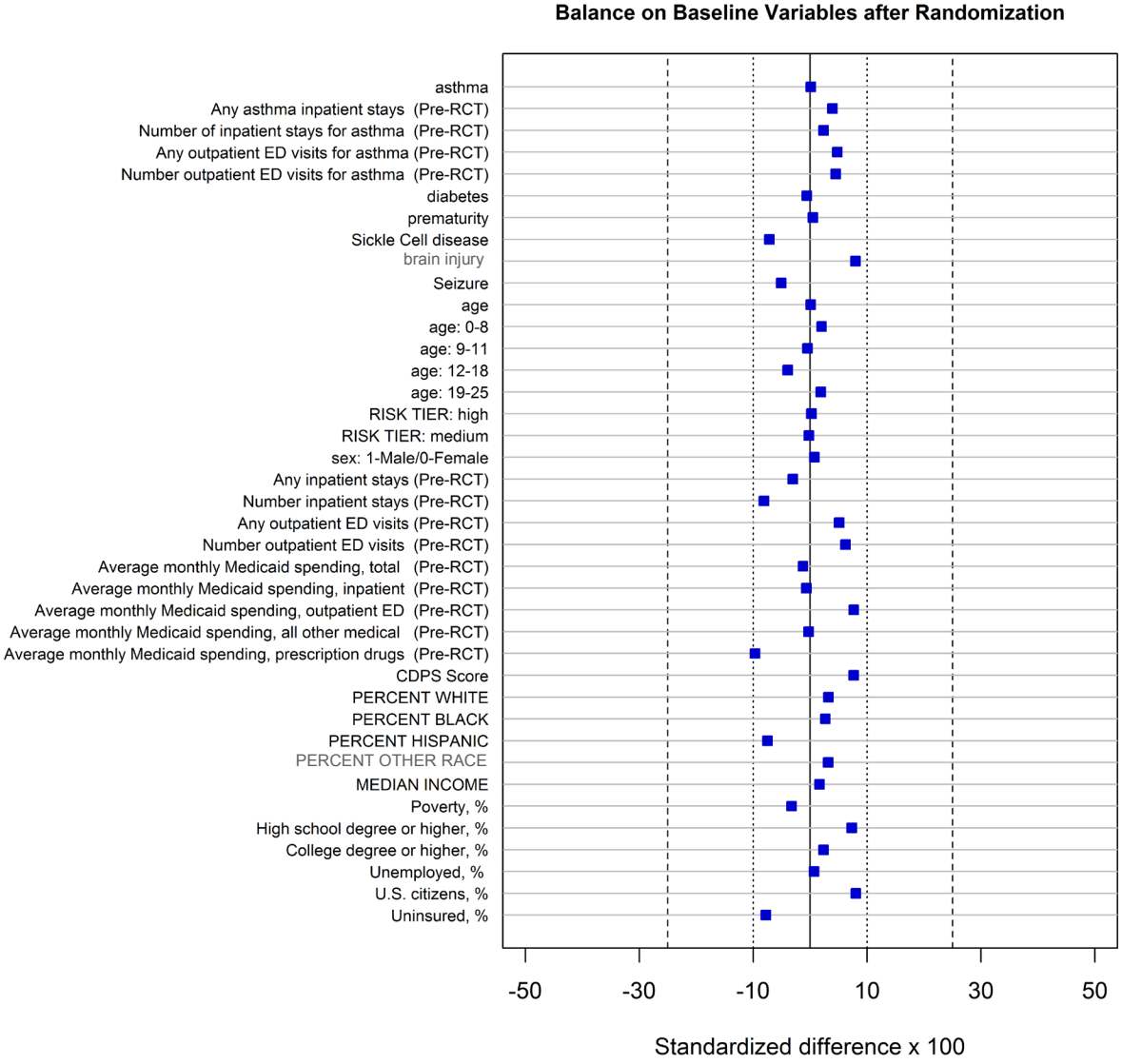


Figure III.2. Standardized differences between treatment and control beneficiaries on baseline variables: subgroup of medium- and high-risk beneficiaries, University of Illinois



Notes: The medium- and high-risk subgroup comprises patients who had one to three emergency department (ED) visits or one hospitalization in the year before randomization (medium risk) or more than three ED visits or more than one hospitalization in the year before randomization (high-risk).

E. Interim impact findings

This section presents preliminary impact estimates over the first year of the RCT period, April 1, 2016, through March 31, 2017. In addition to reporting annual impacts for this first year, we also present semiannual impact estimates in the appendix. These semiannual impact estimates separately cover months 1 to 6 and months 7 to 12 after the start of the RCT. The rationale for assessing semiannual impacts in addition to our primary, annual analyses is to determine whether the program may have had differential impacts over time. For example, if it took the program

several months to fully gain the cooperation of beneficiaries and their families, we might not expect to see impacts in the first half of the year, but perhaps would begin to see impacts in the second half of the year. If this were the case and we evaluated only annual impacts, we could miss an indication that the program was starting to improve outcomes during the second part of the year.

1. Data and estimation approach

The University of Illinois analysis file was constructed using data from the randomization file (demographic characteristics, target conditions, secondary conditions, and risk tier), Medicaid claims data (health care utilization and spending), and zip-code–level data from the American Community Survey. As noted, we fit regression models for the full sample for the first year of the RCT. We also fit regression models for months 1 to 6 and months 7 to 12 to assess whether there were different impacts at different points during the year. We repeated these analyses on the subgroup of beneficiaries who were flagged as medium or high risk ahead of randomization to assess whether the program had greater impacts on beneficiaries with higher risk than it did on the full population. The impact methodology (included as Appendix B to the main body of this report) provides more detailed descriptions of the regression models used for the University of Illinois.

All regression models adjusted for the following covariates: a series of interaction terms between age category (birth to 8, 9 to 11, 12 to 18, and 19 to 25) and gender, indicator variables for whether the beneficiary had any of the four target conditions (asthma, diabetes, prematurity, and sickle cell disease) or any of the two secondary conditions (newborn brain injury or epilepsy/seizure), categorical variable for risk tier (high, medium, or low), CDPS score from the baseline period, and variables for the number of hospitalizations in the baseline period, number of outpatient ED visits in the baseline period, and average monthly Medicaid spending in the pre-RCT year. We also adjusted for zip-code–level variables for the percentage of the zip code that was white, black, Hispanic, or some other race or ethnicity.

Most beneficiaries in the RCT were enrolled in managed care during the baseline and RCT periods. However, some beneficiaries switched to FFS over the analysis period. Medicaid spending in our analyses reflects the sum of all FFS spending and all *estimated* managed care spending. To clarify, the payment variable on the managed care encounter claims reflects what the FFS payment would have been for that service, not necessarily what the managed care organization actually paid (Care Coordination Claims Data [CCCD] data dictionary 2015). Thus, our estimates of Medicaid spending may differ from actual payment amounts.

2. Outcomes

Table III.4 lists the outcomes measures developed for the CHECK evaluation. We assessed changes in the total number and probability of all-cause hospitalizations and outpatient ED visits. We used the PBPM measure to assess total spending and specific spending on inpatient hospital care, outpatient ED visits, prescription drugs, and all other Medicaid-covered services. Because the vast majority of treatment and control beneficiaries have asthma, we also assessed whether the intervention impacted asthma-related hospitalizations and outpatient ED visits. We determined that a hospitalization or an outpatient ED visit was asthma-related if the primary diagnosis on the claim was for asthma.

Table III.4. Outcome measures

Outcome
Average Medicaid expenditures PBPM (FFS and managed care)
Total
Hospitalizations
Outpatient ED visits
Prescription drugs
All other Medicaid-covered services
Health care utilization rates (annualized per 1,000)
Hospitalizations
Outpatient ED visits
Asthma-related hospitalizations
Asthma-related outpatient ED visits
Measures of any health care utilization
Hospitalizations
Outpatient ED visits
Asthma-related hospital admissions
Asthma-related outpatient ED visits

PBPM = per beneficiary per month; FFS = fee-for-service; ED = emergency department.

3. Impact estimates: full RCT population

In this section, we highlight impact estimates that were statistically significant at the 0.10-level over the first full 12-month period of the RCT for the full RCT population. For significant estimates, we provide results at the 90 percent confidence interval (90% CI) to give the reader more information on the estimated range where the true impacts are likely to be found. We note in the text where results might differ for semiannual analyses and refer readers to appendix tables when relevant.

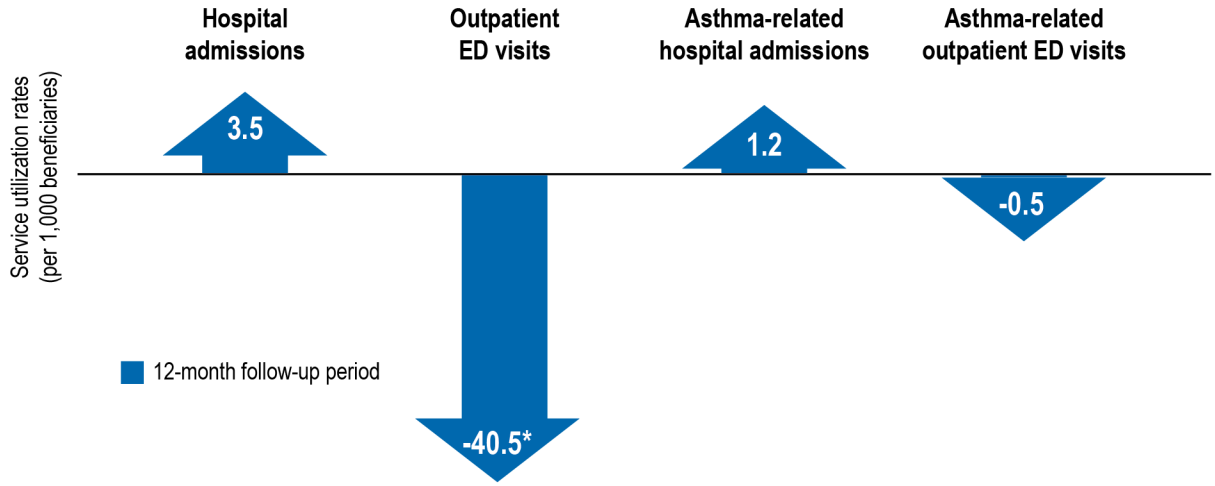
Expenditures

In the full population, there was no observed reduction in total Medicaid spending or in any subcategory of Medicaid spending in the first year of the RCT.

Service utilization

In the full population, the impact estimate for outpatient ED visits suggests that the treatment group had 41 fewer visits per 1,000 beneficiaries than the control group did during the first year of the RCT (90% CI = [-79 per 1000 beneficiaries, -2 per 1,1000 beneficiaries]). The treatment group also had a lower likelihood of any outpatient ED visit over this first year. The share with any outpatient ED visit during the first year of the RCT was 29.6 percent for the treatment group versus 31.9 percent for the control group, a 2.4 percentage point difference (90% CI = [-4.2, -0.6]). When we assessed changes in service utilization over months 1 to 6 and 7 to 12, we found that the treatment group had lower rates of outpatient ED visits in both periods than the control group did, but the differences were larger and statistically significant at the 0.10 level in months 1 to 6 only (Appendix Table B.2). This suggests that most of the reduction in outpatient ED visits took place within the first six months of the RCT period.

Figure III.3. Estimated impact on service utilization rates: full population, University of Illinois



Source: Mathematica analysis of information from randomization file and Medicaid claims and encounter data between April 1, 2016, and March 31, 2017.

Note: Number of treatment beneficiaries = 3,131; number of control beneficiaries = 3,128.

*Significantly different from zero at the .10 level, two-tailed test.

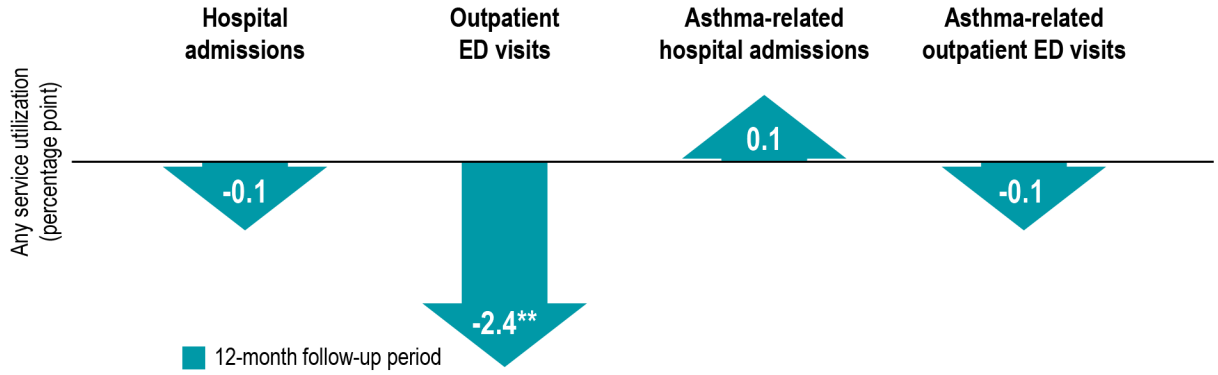
**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

PBPM = per member per month.

There were no differences between the treatment and control groups in the number of all-cause hospitalizations or the probability of any hospitalizations over the first full year of the RCT. However, in months 7 to 12, the probability of any hospitalization was 1.4 percent for the treatment group versus 2 percent for the control group, a 0.6 percentage point difference (90% CI = [-1.1, -0.1]) (Appendix Table B.3). We found no differences between the treatment and control beneficiaries in asthma-related hospitalizations or outpatient ED visits over the full year or in any semiannual period for the full group (Appendix Tables B.1–B.3).

Figure III.4. Estimated impact on likelihood of any service utilization: full population, University of Illinois



Source: Mathematica analysis of information from randomization file and Medicaid claims and encounter data between April 1, 2016, and March 31, 2017.

Note: Number of treatment beneficiaries = 3,131; number of control beneficiaries = 3,128.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

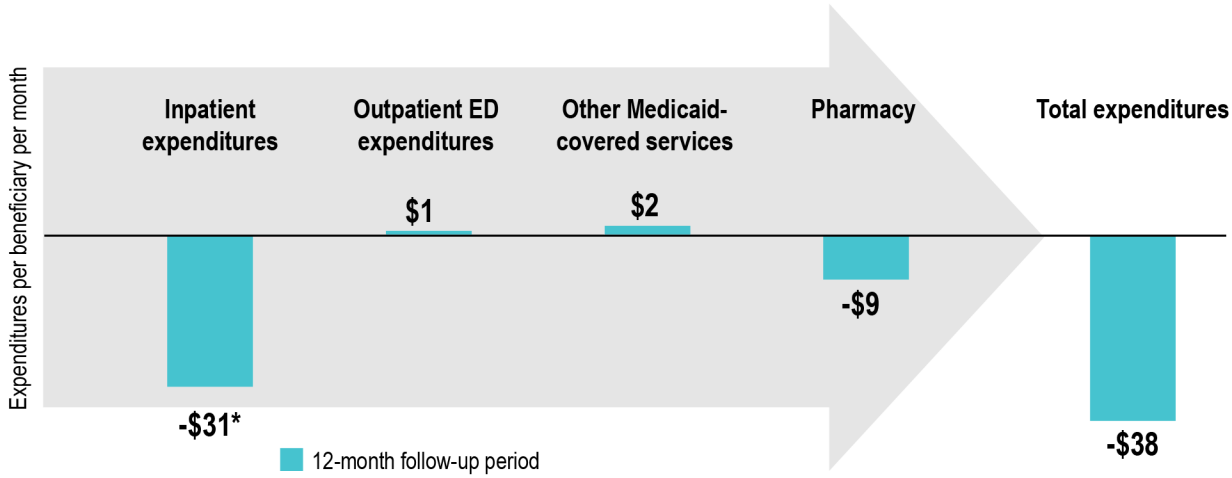
ED = emergency department.

4. Medium- and high-risk subgroup

Expenditures

In the subgroup of medium- and high-risk beneficiaries, average PBPM *inpatient* spending was estimated to be $31 lower (90% CI = [-$60, -$2]) for the treatment group than the control group over the full 12-month period. In semiannual analyses, average *total* Medicaid PBPM spending was $50 lower (90% CI = [-$93, -$6]) for the treatment group than the control group in months 7 to 12. Average *inpatient* PBPM spending in months 7 to 12 was estimated to be $31 lower (90% CI = [-$56, -$6]) for the treatment group than for the control group (Appendix Table B.6).

Figure III.5. Estimated PBPM impact on total expenditures and by expenditure category: subgroup of medium- and high-risk beneficiaries, University of Illinois



Source: Mathematica analysis of information from randomization file and Medicaid claims and encounter data between April 1, 2016, and March 31, 2017.

Note: Number of treatment beneficiaries = 3,131; number of control beneficiaries = 3,128.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

ED = emergency department.

Service utilization

We found no statistically significant reductions in hospitalizations or outpatient ED visits over the full 12 months of the RCT period for the subgroup. However, we found that the treatment group had 30 fewer hospitalizations per 1,000 beneficiary months (90% CI = [-55 per 1,000 beneficiary months, -4 per 1,000 beneficiary months]) than the control group in months 7 to 12. The treatment group also had a lower likelihood of any hospitalization than the control group did in months 7 to 12. The share with any hospitalization during months 7 to 12 was 2 percent for the treatment group versus 5 percent for the control group, a 3 percentage point difference (90% CI = [-4.5, -1.5]). Further, the medium- and high-risk subgroup also had a lower likelihood of any outpatient ED visit in months 7 to 12 compared to the control group. The share with any outpatient ED visit was 20.4 percent for the treatment group and 24.3 percent for the control group, a 4 percentage point difference (90% CI = [-7.5,-0.5]) (Appendix Table B.6). There were no differences between treatment and control subgroups in asthma-related hospitalizations or outpatient ED visits over the full year or any semiannual period (Appendix Tables B.4 –B.6).

5. Sensitivity analyses

For both count and continuous outcomes, we winsorized (that is, truncated) the data at the 99th percentile in sensitivity analyses to test whether outliers may be driving results. Our

conclusions remained the same when we used the winsorized data, although for spending outcomes in the subgroup analyses, the treatment-control differences became much smaller. For example, the impact estimate for total Medicaid spending in our primary analyses of the medium- and high-risk subgroup suggested that average *inpatient* PBPM spending over the full first year of the RCT was $31 lower (90% CI = [-$60, -$2]) for the treatment group than the control group. Using the winsorized data, the impact estimate on *inpatient* spending suggested that it was only $8 lower (90% CI = [-$16, -$0.06]) than the control group over that period.

6. Discussion: The University of Illinois

Our findings suggest that the University of Illinois program, CHECK, reduced outpatient ED visits in the full population in the first year of the RCT period. Most of that reduction took place within the first six months of the RCT. In addition, CHECK reduced Medicaid *inpatient* expenditures for the medium- and high-risk subgroup in the first year of the RCT. The reduction in inpatient expenditures was driven by reductions in both any inpatient hospital care and in the number of hospitalizations among the treatment group during the second half of the first RCT year. We also found evidence of reduced total Medicaid spending per beneficiary per month and on any outpatient ED visits for the medium- and high-risk subgroup in months 7 to 12.

Our findings are based on a randomized controlled trial, the most rigorous study design. There were two factors that may have helped CHECK demonstrate impacts early in the RCT period. First, the RCT started in April 2016, more than 15 months after CHECK began. Thus, the RCT period did not overlap with the first year of the intervention, when there was less infrastructure, fewer resources, and more confusion on the part of staff about their roles and responsibilities. If the RCT had started at the same time as the intervention, we might not have seen impacts during the first year. Second, most of the treatment group beneficiaries who were engaged or activated in the first RCT year were engaged and activated in its first quarter. This gave CHECK more time to demonstrate impacts than the program would have had if the process of activating these beneficiaries had been slower.

CHECK demonstrated impacts, even though only one-quarter of treatment group beneficiaries were engaged or activated. The extent to which impacts would have been larger if all treatment group beneficiaries had been engaged is unknown. It may be that CHECK successfully reached all those beneficiaries and families who were willing to participate and had the greatest need for help, and that our current impact estimates reflect the actual impacts. Alternatively, there may be other beneficiaries and families in the treatment group for whom CHECK could have helped improve outcomes if the awardee had been able to engage or activate the beneficiaries and families. In this case, our impact estimates might understate the potential impacts of the intervention.

The relatively low rate of engagement or activation among the treatment group likely reflects some of the problems inherent in relying on data from managed care organizations to facilitate initial contact with participants. As noted, program leaders and frontline staff reported that the contact information the program received from managed care plans was often out of date or incomplete, making it difficult for program staff to find and communicate with participants who were enrolled in the program (that is, individuals who were identified as eligible and to whom the program mailed an enrollment letter). Program leaders noted that contact information

from the managed care organization that participated in the impact evaluation may have posed particular challenges.

The relatively low rates of engagement and activation may also reflect issues that arise in working with a low-income population. It may be hard to contact beneficiaries (even those with initially valid contact information), if they move often or their phone numbers change. These beneficiaries may lack trust in the health care system. This particular intervention, moreover, was focused on children and young adults, meaning the program needed to engage their caregivers. The caregivers themselves may have had their own medical or behavioral issues which, despite the program's efforts to work with participants' families and caregivers, could have kept them from getting engaged in CHECK. The caregivers also may have had limited time available to work with the community health workers.

The rapid reduction in outpatient ED visits among treatment group beneficiaries in the full sample suggests that CHECK quickly helped connect families and beneficiaries to needed health and community-based resources. The reduction in inpatient spending and associated utilization in the second half of the year for the medium- and high-risk beneficiaries suggests that the program needed more time to demonstrate that it reduced hospitalizations, and that this reduction was concentrated among the sickest patients. We do not have any data on the number of contacts per engaged or activated beneficiary—or data that would characterize the types of assistance provided to engaged and activated beneficiaries. Such data would help us understand the correlation between the program and outcomes.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

At the end of the third year of its cooperative agreement, University of Illinois proposed a PBPM care coordination fee for Medicaid MCOs as its primary payment model, but direct negotiations were on hold pending the re-bidding of MCO contracts by the state Medicaid agency in the fall of 2017. The awardee was also exploring a direct employment relationship with its home health system, but the system had little interest in a program focused on children.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: (1) the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor and (2) narratives and payment model reports the awardee submitted to the implementation and monitoring contractor. We did not interview the awardee about its payment model because it did not advance from the development stage to active negotiations with one or more payers.

B. Description of the payment model

The primary payment model proposed by the University of Illinois was a PBPM fee that a Medicaid MCO would pay to the program for each enrolled beneficiary. The fee would cover the same services implemented under the award for the same target populations. With support from an actuarial consultant, the awardee estimated the PBPM would depend on the number of beneficiaries enrolled, and range from $23.13 for about 14,000 beneficiaries to $55.31 for about 6,000 beneficiaries. The awardee based these estimates on analyses of its fixed labor, fixed non-labor, and variable costs over the first two years of the cooperative agreement. However, at the end of the third program year, the awardee did not yet have analyses of the cost savings the program realized for participants, and therefore could not perform a full return-on-investment analysis. The awardee did not propose any adjustments to the fees based on beneficiary complexity, quality measures, or spending benchmarks. The awardee proposed focusing on the same populations and delivering the same services that it did under the cooperative agreement, but acknowledged the need to negotiate these details with payers.

C. Status of the payment model

Near the end of the third year of the cooperative agreement, University of Illinois stated that its direct negotiations with payers were on hold as the state Medicaid agency re-bid its Medicaid managed care contracts. The awardee planned to resume negotiations with its prior partner MCOs or any new MCOs after the expected awarding of contracts in January 2018. The awardee planned to repeat its analysis of program costs including a full three years of data and to re-estimate PBPM rates as a starting point for negotiations. Awardee leaders also planned more analyses to determine the labor costs for specific staff types based on tasks involved with specific patient conditions, to estimate the impact of the program on health care utilization and spending, and to assess associations between spending and specific program services. They planned to use these analyses in negotiations with potential payers.

In addition to the PBPM fee payment model, the awardee was also exploring opportunities for direct employment or contracting with its home or other health systems.

D. Factors associated with the development of the payment model

The awardee's leaders described three significant barriers to the development of a payment model for the program: (1) the state Medicaid agency's implementation and re-bidding of managed care, (2) the difficulty of obtaining and using Medicaid claims data, and (3) limited interest in care coordination from its home institution.

First, the Illinois Medicaid agency began implementing Medicaid managed care at the same time the CHECK program started, which required the awardee to shift from plans to negotiate for access to data and develop a payment model with the Medicaid agency to working with multiple new MCOs. Although the awardee was able to build partnerships for data exchange with two MCOs, the state was in the process of rebidding Medicaid managed care contracts near the end of the third year of the cooperative agreement, which put the awardee's plans to begin negotiations with the MCOs on hold.

Second, the awardee faced challenges throughout the award in obtaining complete claims data from the state Medicaid agency or MCOs, limiting its ability to estimate the program's impacts on health care utilization.

Third, when awardee leaders discussed the possibility of ongoing institutional support for the program, they found that health system leaders had little interest in a program focused on children. To help address this, they had begun discussions about the possibility of piloting the program with adults, based on their experience working with parents of enrolled children.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The University of Illinois reported that it was working to analyze program data to both build a business case for the program that would engage payers and organizational leaders, and to determine the most efficient staffing models and operational budgets for the program. The awardee also reported scaling activities in the third program year, including pilot programs in community-based health centers and an expanded presence in the university's hospital. The awardee did not report plans to replicate its program.

The biggest challenge to sustainment and scaling reported by the awardee was inadequate and unreliable funding sources for care coordination programs for children. It worked to overcome this challenge by diversifying the program's funding sources.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, the University of Illinois had begun work on its sustainability plan to engage MCOs to pay a PBPM fee for program services. The awardee had started working with Milliman, a health actuarial firm, to calculate the PBPM fee, and planned to market its care coordination program to MCOs as a flexible approach to non-clinical care

coordination that allowed the MCOs to adapt the program to their patient population by selecting available resources from a menu of program services. The awardee also reported a potential strategy of integrating program services with hospital discharge planning at the university's hospital. Finally, the awardee reported that it was broadening its "medical neighborhood" by building relationships with community organizations and health clinics.

C. Implementing the SSR plan: progress and changes

Sustainability. According to awardee leaders, changes made to the program to facilitate expansion and sustainment will be based on requirements by the funding source, as well as community- and region-specific needs. As of the third program year, the University of Illinois focused on a variety of sustainment strategies.

First, awardee leaders described their efforts to analyze program data and build a business case for the program. They worked with a team of marketing students from the University of Chicago to develop a program pitch and different ways to present program data to different audiences—payers versus organizational leaders, for example.

Second, CHECK program leaders worked to sustain the program as a potentially permanent unit in the health system. According to awardee leaders, these efforts included analyses of program data by CHECK's technology and data team to determine ways to improve the program. Using program data and input from stakeholders, the University of Illinois developed a staffing model and operational budget to support program sustainment. The awardee also sought to grow the program presence within the university hospital, engaging with departmental leaders in pediatrics, general medicine, emergency medicine, and psychiatry to further integrate the program into the hospital's care approach.

Third, the university's hospital hired an administrative lead for population management for the pediatric and internal medicine departments in the third program year; site visit interviewees believed this person would collaborate with the program.

Fourth, the awardee focused on sustaining partnerships with important stakeholders in the community. The awardee's marketing team developed marketing materials to distribute to key stakeholders, including community organizations, elected officials, churches, and other partnering community health sites.

Scalability. The University of Illinois is also looking at scaling variations of the program. First, the awardee reported plans to start a related pilot program in seven sites of a federally qualified health center owned by the University of Illinois. Second, the awardee reported plans to lease program technology to interested organizations. The awardee received six months of funding from the Chicago Innovators Mentorship to develop a public-private partnership with an independent company to license all software products developed for the CHECK program for use by other programs. After that funding ended, the awardee planned to apply for other funding to finish developing the technology.

Replicability. The University of Illinois did not report any plans to replicate its program in the third program year.

D. Factors associated with progress toward implementing the SSR plan

The University of Illinois reported funding as the biggest challenge to sustaining the program. Awardee leaders said that uncertainty in state politics made it difficult for program leaders to rely on any one source of public funding. Instead, the awardee began looking into a variety of sources, including external grants and revenue from leasing the program's technology, in addition to engaging payers. Awardee leaders explained that "child health dollars are a relatively small component of the health care system," making it difficult to find adequate levels of funding despite a large demand for managed care programs for children. Regardless, awardee leaders expressed hope that a growing number of payers would be interested in funding the program, and expect that as insurance companies become accountable for quality and health outcomes, they may increasingly rely on low-cost care coordination programs to support broader population health management.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension that allows it to continue to enroll new beneficiaries. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

In 2018, we plan to assess impacts of the CHECK program over the second year of the RCT (April 2017–March 2018). We will talk to the University of Illinois about whether there might be data available on the number and types of contacts available for engaged and activated treatment group members that we might use in next year's analyses.

In addition, our analyses suggested that few beneficiaries had asthma-related hospitalizations or outpatient ED visits, and there were no impacts on these measures of utilization. The low rate of asthma-related utilization may be because of the strict definition we applied. Specifically, we only flagged utilization as asthma-related if the primary diagnosis on the claim was for asthma. The rationale for limiting asthma utilization to claims with a primary diagnosis was to test whether utilization caused specifically by asthma declined. However, it might be preferable to expand the definition of asthma-related utilization to include secondary diagnoses to examine whether effective treatment of asthma might be a contributing factor in reductions in hospitalizations and ED visits. For example, it is possible that without the intervention, children with asthma who caught pneumonia or a bad case of the flu would have been hospitalized, but because of the intervention, the parents were better able to manage the asthma and prevented the ED visit or hospital stay. Thus, for next year's report, we plan to update the definition of asthma-related for the impacts analyses to use primary and secondary diagnoses.

We did not construct a measure for pediatric 30-day unplanned readmissions for this annual report. Relatively few beneficiaries in the full sample had any hospitalizations, making it hard to detect impacts on readmissions at the population level. As an alternative, we could try to assess readmissions only among those who had a hospitalization. The problem with this approach is that to the extent the intervention reduces hospitalizations for the treatment group, those that do occur may be for higher-risk and sicker beneficiaries whose stays might be more likely to be

followed by a readmission. Thus, we might observe an increased readmission rate for the treatment group in this situation. As a result, we do not recommend pursuing an analysis of this outcome measure.

REFERENCES

Care Coordination Claims Data (CCCD) Data Dictionary. Last updated November 1, 2015. Provided to Mathematica Policy Research by CHECK program.

Centers for Disease Control and Prevention. "Most Recent Asthma Data." Last updated June 2017a. Available at https://www.cdc.gov/asthma/most_recent_data.htm. Accessed December 9, 2017.

Centers for Disease Control and Prevention. "National Diabetes Statistics Report, 2017." Atlanta, GA: Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, 2017b. Available at https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf. Accessed December 9, 2017.

Centers for Disease Control and Prevention. "More Americans Have Epilepsy than Ever Before." Last updated August 2017c. Available at https://www.cdc.gov/media/releases/2017/p0810-epilepsy-prevalence.html. Accessed December 9, 2017.

Kronick, R., T. Gilmer, T. Dreyfus, and L. Lee. "Improving Health-Based Payments for Medicaid Beneficiaries: CDPS." *Health Care Financing Review,* vol. 21, no. 3, 2000, pp. 29–64. Available at http://cdps.ucsd.edu/cdps_hcfr.pdf. Accessed January 10, 2016.

National Heart, Lung, and Blood Institute. "Who Is at Risk for Sickle Cell Disease?" Last updated August 2017. Available at https://www.nhlbi.nih.gov/health/health-topics/topics/sca/atrisk. Accessed December 9, 2017.

This page has been left blank for double-sided copying.

APPENDIX A

POST-MATCHING DIAGNOSTICS

This page has been left blank for double-sided copying.

Table A.1. Balance between treatment and control group on key characteristics: full sample, University of Illinois

Measure	Control (n = 3128)	Treatment (n = 3131)	Difference	Percentage difference	Standardized difference	*p*-value	Equivalence *p*-value
Age and gender							
Age at enrollment (continuous)	11.2	11.2	-0.041	-0.400	-0.006	0.799	0.000
Age group, %							
0–8 years	39	39	0.002	0.500	0.004	0.880	0.000
9–11 years	15	16	0.003	2.000	0.008	0.738	0.000
12–18 years	31	31	-0.003	-0.800	-0.005	0.829	0.000
19–25 years	15	15	-0.002	-1.600	-0.007	0.791	0.000
Male, %	53	53	0.001	0.200	0.002	0.931	0.000
Target conditions, %							
Asthma	87	86	-0.007	-0.800	-0.019	0.448	0.000
Diabetes	6	6	0.000	-0.100	0.000	0.993	0.000
Prematurity	5	5	0.000	0.600	0.001	0.959	0.000
Sickle cell disease	1	1	-0.001	-19.200	-0.016	0.552	0.000
Secondary conditions, %							
Newborn brain jury	0.2	0.4	0.002	50.000	0.033	0.127	0.000
Epilepsy/seizure	4	4	0.001	2.200	0.005	0.855	0.000
Risk tier, %							
High	4	4	0.000	0.600	0.001	0.957	0.000
Medium	22	22	0.000	0.100	0.000	0.992	0.000
Low	74	74	-0.001	-0.100	-0.002	0.949	0.000
Health status and Medicaid utilization and expenditures pre-RCT							
CDPS score	2.5	2.7	0.151	5.600	0.041	0.073	0.000
Any hospitalizations, %	6	5	-0.005	-8.500	-0.020	0.435	0.000
Number of hospitalizations	0.09	0.07	-0.016	-22.500	-0.043	0.143	0.000
Any outpatient ED visits, %	33	34	0.006	1.700	0.012	0.629	0.000
Number of outpatient ED visits	0.59	0.62	0.021	3.500	0.017	0.485	0.000
Any asthma hospitalizations, %	1	1	0.002	23.700	0.020	0.411	0.000

Table A.1 *(continued)*

Measure	Control (n = 3128)	Treatment (n = 3131)	Difference	Percentage difference	Standardized difference	*p*-value	Equivalence *p*-value
Number of hospitalizations for asthma	0.01	0.01	0.001	14.200	0.012	0.648	0.000
Any outpatient ED visits for asthma, %	4	5	0.001	2.000	0.004	0.861	0.000
Number outpatient ED visits for asthma	0.06	0.06	0.008	12.200	0.017	0.430	0.000
Total Medicaid expenditures ($ PBPM)	$203	$183	-19.883	-10.900	-0.016	0.564	0.000
Inpatient expenditures ($ PBPM)	$83	$71	-11.825	-16.600	-0.010	0.712	0.000
Outpatient ED expenditures ($ PBPM)	$9	$9	0.018	0.200	0.001	0.979	0.000
All other medical expenditures ($ PBPM)	$77	$73	-3.840	-5.300	-0.027	0.417	0.000
Prescription drug expenditures ($ PBPM)	$34	$30	-4.236	-14.200	-0.041	0.172	0.000
Zip-code–level characteristics							
White, %	25	26	0.157	0.600	0.006	0.802	0.000
Black, %	38	38	0.164	0.400	0.005	0.851	0.000
Hispanic, %	33	32	-0.459	-1.400	-0.017	0.495	0.000
Other race/ethnicity, %	4	5	0.138	3.000	0.027	0.278	0.000
Median income	$45,812	$45,810	-2.462	0.000	0.000	0.995	0.000
Poverty, %	23	23	-0.105	-0.500	-0.010	0.702	0.000
High school degree or higher, %	77	77	0.210	0.300	0.017	0.494	0.000
College degree or higher, %	20	20	0.045	0.200	0.004	0.886	0.000
Unemployed, %	15	15	0.004	0.000	0.001	0.980	0.000
U.S. citizens, %	89	89	0.114	0.100	0.012	0.627	0.000
Uninsured, %	18	17	-0.098	-0.600	-0.016	0.527	0.000

Notes: Difference calculated as the ratio of the difference and the treatment group standard deviation. *p*-values come from a two-sample t-test; equivalence test *p*-values are the greater of two one-sided t-test *p*-values for a test of whether the treatment and control group means differ by more than .1 standard deviations.

PBPM = per beneficiary per month; FFS = fee-for-service; ED = emergency department; CDPS = Chronic Illness and Disability Payment System.

Table A.2. Balance between treatment and control groups on key characteristics: subgroup of medium-high risk beneficiaries, University of Illinois

Measure	Control (n = 819)	Treatment (n= 821)	Difference	Percentage Difference	Standardized difference	*p*-value	Equivalence *p*-value
Age and gender							
Age at enrollment (continuous)	11.5	11.5	0.006	0.100	0.001	0.987	0.000
Age group, %							
0–8 years	42	43	0.010	2.300	0.020	0.684	0.000
9–11 years	11	11	-0.001	-1.400	-0.005	0.923	0.000
12–18 years	24	22	-0.016	-7.400	-0.040	0.429	0.000
19–25 years	23	24	0.008	3.300	0.019	0.704	0.000
Male, %	49	50	0.004	0.700	0.007	0.882	0.000
Target conditions, %							
Asthma	84	84	0.000	0.000	0.001	0.983	0.000
Diabetes	6	6	-0.001	-2.300	-0.006	0.908	0.000
Prematurity	6	6	0.001	1.800	0.005	0.926	0.000
Sickle cell disease	1	1	-0.006	-83.800	-0.072	0.221	0.001
Secondary conditions, %							
Newborn brain injury	0.1	1	0.007	85.700	0.079	0.034	0.000
Epilepsy/seizure	6	5	-0.011	-22.200	-0.051	0.326	0.000
Risk tier, %							
High	0.17	0.17	0.001	0.500	0.002	0.965	0.000
Medium	0.83	0.83	-0.001	-0.100	-0.002	0.965	0.000
Health status and Medicaid utilization and expenditures pre-RCT							
CDPS score	3.3	3.6	0.357	9.800	0.076	0.071	0.000
Any hospitalizations, %	18	17	-0.011	-6.700	-0.030	0.545	0.000
Number of hospitalizations	0.28	0.23	-0.052	-22.800	-0.081	0.170	0.002
Any outpatient ED visits, %	71	73	0.023	3.100	0.051	0.307	0.000
Number of outpatient ED visits	1.5	1.6	0.119	7.400	0.062	0.178	0.000
Any asthma hospitalizations, %	2	2	0.006	24.800	0.039	0.397	0.000
Number of hospitalizations for asthma	0.02	0.02	0.004	14.800	0.023	0.641	0.000

Table A.2 *(continued)*

Measure	Control (n = 819)	Treatment (n= 821)	Difference	Percentage Difference	Standardized difference	*p*-value	Equivalence *p*-value
Any outpatient ED visits for asthma, %	11	12	0.016	12.800	0.048	0.321	0.000
Number outpatient ED visits for asthma	0.15	0.18	0.037	20.500	0.045	0.269	0.000
Total Medicaid expenditures, ($ PBPM)	$391	$372	-18.816	-5.100	-0.013	0.807	0.000
Acute inpatient expenditures ($ PBPM)	$200	$191	-9.463	-5.000	-0.007	0.892	0.000
Outpatient ED expenditures ($ PBPM)	$22	$26	3.626	14.100	0.076	0.069	0.000
All other medical expenditures ($ PBPM)	$117	$117	-0.579	-0.500	-0.003	0.956	0.000
Prescription drug expenditures ($ PBPM)	$52	$39	-12.400	-31.500	-0.097	0.179	0.020
Zip-code–level characteristics							
White, %	29	30	0.866	2.900	0.032	0.515	0.000
Black, %	34	35	0.905	2.600	0.027	0.588	0.000
Hispanic, %	32	30	-1.934	-6.400	-0.075	0.132	0.000
Other race/ethnicity, %	5	5	0.164	3.300	0.032	0.512	0.000
Median income	$47,765	$48,069	304.471	0.600	0.016	0.741	0.000
Poverty, %	22	22	-0.361	-1.600	-0.033	0.504	0.000
High school degree or higher, %	77	78	0.928	1.200	0.073	0.145	0.000
College degree or higher, %	20	21	0.308	1.500	0.024	0.639	0.000
Unemployed, %	14	14	0.044	0.300	0.007	0.890	0.000
U.S. citizens, %	89	89	0.733	0.800	0.080	0.109	0.000
Uninsured, %	17	17	-0.506	-3.100	-0.078	0.115	0.000

Notes: Standard errors in parentheses. Standardized difference calculated as the difference in means divided by the treatment group standard deviation. *The p*-values come from a weighted two-sample t-test. The equivalence test comprises two one-sided tests of whether the difference between the treatment and control group means exceeds 0.25 standard deviations (in either direction) of that variable. If we can reject that hypothesis, we are confident that the true difference between the two populations being compared is less than 0.25 standard deviations. The medium- and high-risk subgroup comprises patients who had one to three emergency department (ED) visits or one hospitalization in the year before randomization (medium risk) or more than three ED visits or more than one hospitalization in the year before randomization (high-risk).

PBPM = per beneficiary per month; FFS = fee-for-service; ED = emergency department; CDPS = Chronic Illness and Disability Payment System.

APPENDIX B

IMPACT ESTIMATE TABLES

This page has been left blank for double-sided copying.

Table B.1. Estimated impact of the University of Illinois program on Medicaid FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a 12-month follow-up period: full population

Expenditures and utilization measures	Treatment (n = 3,131)	Control (n = 3,128)	Impact (Treatment – Control)	*p*-value	90% confidence interval
Average Medicaid expenditures PBPM ($)[a]					
Total	119	130	-11	0.293	(-29, 6)
Inpatient	24	28	-4	0.649	(-18, 10)
Outpatient ED	7	8	0	0.647	(-1, 1)
Other Medicaid-covered services	57	61	-4	0.235	(-10, 2)
Pharmacy	30	33	-3	0.400	(-8, 3)
Health care utilization rates (annualized per 1,000)					
Hospital admissions	50	47	4	0.659	(-9, 16)
Outpatient ED visits	482	523	-41*	0.083	(-78, -2)
Asthma-related hospitalizations[b]	3	2	1	0.354	(-1, 3)
Asthma-related outpatient ED visits[b]	37	38	-1	0.934	(-10, 9)
Measures of any health care utilization (%)					
Percentage with a hospital admission[d]	3.6	3.7	-0.1	0.816	(-0.9, 0.7)
Percentage with an outpatient ED visit	29.6	31.9	-2.4**	0.039	(-4.2, -0.6)
Percentage with an asthma-related hospital admission[b]	0.3	0.2	0.1	0.354	(-0.1, 0.3)
Percentage with an asthma-related outpatient ED visit[b]	3.1	3.2	-0.1	0.791	(-0.9, 0.7)

Source: Mathematica analysis of information from awardee's randomization file and Medicaid claims and encounter data between April 1, 2016, and March 31, 2017.

[a]Total Medicaid expenditures were calculated from all claims for each participant with at least one eligible day during the intervention period.

[b]Asthma-related utilization as defined as any inpatient hospital stay or outpatient ED visit with an asthma primary diagnosis.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

PBPM = per beneficiary per month; ED = emergency department.

Table B.2. Estimated impact of the University of Illinois program on Medicaid FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a 1–6 month follow-up period: full population

Expenditures and utilization measures	Treatment (n = 3,131)	Control (n= 3,128)	Impact (Treatment – Control)	*p*-value	90% confidence interval
Average Medicaid expenditures PBPM ($)[a]					
Total	123	143	-20	0.118	(-40, 1)
Inpatient	21	34	-13	0.217	(-30, 4)
Outpatient ED	9	9	-1	0.374	(-2, 1)
Other Medicaid-covered services	62	68	-6	0.181	(-13, 1)
Pharmacy	31	31	$0	0.951	(-6, 5)
Health care utilization rates (annualized per 1,000)					
Hospital admissions	30	24	6	0.224	(-2.2, 15)
Outpatient ED visits	284	313	-30*	0.079	(-57, -2)
Asthma-related hospitalizations[b]	2	1	1	0.192	(-0.4, 3.1)
Asthma-related outpatient ED visits[b]	24	22	1	0.747	(-6, 9)
Measures of any health care utilization (%)					
Percentage with a hospital admission[d]	2.4	2.0	0.4	0.244	(-0.3, 1.1)
Percentage with an outpatient ED visit	20.2	21.7	-1.6	0.124	(-3.2, 0.0)
Percentage with an asthma-related hospital admission[b]	0.2	0.1	0.1	0.192	(-0.1, 0.3)
Percentage with an asthma-related outpatient ED visit[b]	1.9	1.9	0.0	0.951	(-0.7, 0.7)

Source: Mathematica analysis of information from awardee's randomization file and Medicaid claims and encounter data between April 1, 2016, and September 30, 2016.

[a]Total Medicaid expenditures were calculated from all claims for each participant with at least one eligible day during the intervention period.

[b]Asthma-related utilization as defined as any inpatient hospital stay or outpatient ED visit with an asthma primary diagnosis.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

PBPM = per beneficiary per month; ED = emergency department.

Table B.3. Estimated impact of the University of Illinois program on Medicaid FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a 7–12 month follow-up period: full population

Expenditures and utilization measures	Treatment (N = 2,937)	Control (N = 2,919)	Impact (Treatment – Control)	*p*-value	90% confidence interval
Average Medicaid expenditures PBPM ($)[a]					
Total	115	117	-2	0.869	(-24, 19)
Inpatient	27	21	6	0.579	(-11, 23)
Outpatient ED	6	6	0	0.699	(-1, 2)
Other Medicaid-covered services	52	54	-3	0.476	(-9, 3)
Pharmacy	30	35	-6	0.185	(-12, $1)
Health care utilization rates (annualized per 1,000)					
Hospital admissions	20	23	-2.9	0.564	(-11, 5.4)
Outpatient ED visits	196	207	-10	0.448	(-32, 12)
Asthma-related hospitalizations[b]	0.7	0.9	-0.2	0.782	(-1.4, 1.0)
Asthma-related outpatient ED visits[b]	14	16	-2.1	0.529	(-7.6, 3.4)
Measures of any health care utilization (%)					
Percentage with a hospital admission[d]	1.4	2	-0.6*	0.071	(-1.1, -0.1)
Percentage with an outpatient ED visit	14.9	15.6	-0.6	0.501	(-2.1, 0.9)
Percentage with an asthma-related hospital admission[b]	0.1	0.1	0.0	0.782	(-0.2, 0.2)
Percentage with an asthma-related outpatient ED visit[b]	1.3	1.5	-0.1	0.649	(-0.6, 0.4)

Source: Mathematica analysis of information from awardee's randomization file and Medicaid claims and encounter data between October 1, 2016, and March 31, 2017.

[a]Total Medicaid expenditures were calculated from all claims for each participant with at least one eligible day during the intervention period.

[b]Asthma-related utilization as defined as any inpatient hospital stay or outpatient ED visit with an asthma primary diagnosis

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

PBPM = per beneficiary per month; ED = emergency department.

Table B.4. Estimated impact of the University of Illinois program on Medicaid FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a 12-month follow-up period: medium- and high-risk subgroup

Expenditures and utilization measures	Treatment (N = 822)	Control (N = 819)	Impact (Treatment – Control)	*p*-value	90% confidence interval
Average Medicaid expenditures PBPM ($)[a]					
Total	158	195	-38	0.136	(-79, 4)
Inpatient	29	60	-31*	0.078	(-60, -2)
Outpatient ED	14	13	1	0.672	(-2, 4)
Other Medicaid-covered services	74	72	2	0.792	(-8, 12)
Pharmacy	41	50	-9	0.395	(-27, 9)
Health care utilization rates (annualized per 1,000)					
Hospital admissions	96	125	-29	0.286	(-74, 16)
Outpatient ED visits	832	897	-65	0.367	(-184, 54)
Asthma-related hospitalizations[b]	5.2	6.5	-1.3	0.731	(-7.7, 5)
Asthma-related outpatient ED visits[b]	65	55	10	0.513	(-15, 35)
Measures of any health care utilization (%)					
Percentage with a hospital admission[d]	6.3	8.4	-2.1	0.112	(-4.2, 0.0)
Percentage with an outpatient ED visit	41.1	44.9	-3.8	0.129	(-7.9, 0.3)
Percentage with an asthma-related hospital admission[b]	0.5	0.6	-0.1	0.731	(-0.8, 0.6)
Percentage with an asthma-related outpatient ED visit[b]	5.3	4.7	0.6	0.581	(-1.2, 2.4)

Source: Mathematica analysis of information from awardee's randomization file and Medicaid claims and encounter data between April 1, 2016, and March 31, 2017.

Notes: The medium- and high-risk subgroup comprises patients who had one to three emergency department (ED) visits or one hospitalization in the year before randomization (medium risk) or more than three ED visits or more than one hospitalization in the year before randomization (high-risk).

[a]Total Medicaid expenditures were calculated from all claims for each participant with at least one eligible day during the intervention period.

[b]Asthma-related utilization as defined as any inpatient hospital stay or outpatient ED visit with a primary diagnosis of asthma.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

PBPM = per beneficiary per month; ED = emergency department.

Table B.5. Estimated impact of the University of Illinois program on Medicaid FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a 1-6 month follow-up period: medium- and high-risk subgroup

Expenditures and utilization measures	Treatment (n = 822)	Control (n = 819)	Impact (Treatment – Control)	*p*-value	90% confidence interval
Average Medicaid expenditures PBPM ($)[a]					
Total	181	207	-26	0.393	(-76, 24)
Inpatient	37	68	-31	0.192	(-70, 8)
Outpatient ED	16	16	0	0.944	(-4, 3)
Other Medicaid-covered services	86	80	6	0.429	(-7, 19)
Pharmacy	41	43	-1	0.886	(-18, 15)
Health care utilization rates (annualized per 1,000)					
Hospital admissions	62	62	-0.1	0.995	(-30, 30)
Outpatient ED visits	504	540	-36	0.456	(-117, 44)
Asthma-related hospitalizations[b]	5	3.8	1.3	0.707	(-4.2, 6.7)
Asthma-related outpatient ED visits[b]	38	28	10	0.356	(-7.9, 28)
Measures of any health care utilization (%)					
Percentage with a hospital admission[d]	4.7	4.3	0.4	0.722	(-1.2, 2.0)
Percentage with an outpatient ED visit	31.0	31.8	-0.8	0.726	(-4.6, 3.0)
Percentage with an asthma-related hospital admission[b]	0.5	0.4	0.1	0.707	(-0.4, 0.6)
Percentage with an asthma-related outpatient ED visit[b]	2.9	2.5	0.4	0.648	(-0.9, 1.7)

Source: Mathematica analysis of information from awardee's randomization file and Medicaid claims and encounter data between April 1, 2016, and September 30, 2016.

Notes: The medium- and high-risk subgroup comprises patients who had one to three emergency department (ED) visits or one hospitalization in the year before randomization (medium risk) or more than three ED visits or more than one hospitalization in the year before randomization (high-risk).

[a]Total Medicaid expenditures were calculated from all claims for each participant with at least one eligible day during the intervention period.

[b]Asthma-related utilization as defined as any inpatient hospital stay or outpatient ED visit with a primary diagnosis of asthma.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

PBPM = per beneficiary per month; ED = emergency department.

Table B.6. Estimated impact of the University of Illinois program on Medicaid FFS expenditures, rates of utilization, and likelihood of any acute care utilization during a 7-12 month follow-up period: medium- and high-risk subgroup

Expenditures and utilization measures	Treatment (n = 785)	Control (n = 775)	Impact (Treatment – Control)	*p*-value	90% confidence interval
Average Medicaid expenditures PBPM ($)[a]					
Total	133	183	-50*	0.060	(-93, -6)
Inpatient	20	51	-31**	0.041	(-56, -6)
Outpatient ED	12	10	2	0.484	(-2, 6)
Other Medicaid-covered services	61	64	-3	0.622	(-14, 7)
Pharmacy	40	58	-17	0.231	(-41, 7)
Health care utilization rates (annualized per 1,000)					
Hospital admissions	33	63	-30*	0.055	(-55, -4)
Outpatient ED visits	325	352	-28	0.481	(-92, 37)
Asthma-related hospitalizations[b]	0	2.7	-2.7	0.157	(-5.8, 0.4)
Asthma-related outpatient ED visits[b]	26	27	-0.3	0.973	(-15, 14)
Measures of any health care utilization (%)					
Percentage with a hospital admission[d]	2.0	4.9	-3.0***	0.002	(-4.5, -1.5)
Percentage with an outpatient ED visit	20.4	24.3	-4.0*	0.065	(-7.5, -0.5)
Percentage with an asthma-related hospital admission[b]	0	0.3	-0.3	0.157	(-0.6, 0.0)
Percentage with an asthma-related outpatient ED visit[b]	2.5	2.5	0.0	0.971	(-1.3, 1.3)

Source: Mathematica analysis of information from awardee's randomization file and Medicaid claims and encounter data between October 1, 2016, and March 31, 2017.

Notes: The medium- and high-risk subgroup comprises patients who had one to three emergency department (ED) visits or one hospitalization in the year before randomization (medium risk) or more than three ED visits or more than one hospitalization in the year before randomization (high risk).

[a]Total Medicaid expenditures were calculated from all claims for each participant with at least one eligible day during the intervention period.

[b]Asthma-related utilization as defined as any inpatient hospital stay or outpatient ED visit with a primary diagnosis of asthma.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

PBPM = per beneficiary per month; ED = emergency department.

This page has been left blank for double-sided copying.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Regents of the University of Michigan

April 18, 2018

Mynti Hossain (Mathematica Policy Research)
Luke Horner (Mathematica Policy Research)
Boyd Gilman (Mathematica Policy Research)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Existing literature on the effect of prehabilitation on surgical outcomes demonstrates early evidence that prehabilitation may provide post-operative physical benefits to patients and reduce

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

length of hospital stay (Mina 2014; Moran 2016).[3] To further test and examine the effects of prehabilitation on surgical outcomes, the University of Michigan began implementing the Michigan Surgical and Health Optimization Program (MSHOP) at the University of Michigan Health System (UMHS), an academic medical center, in 2011. In 2014, the University of Michigan used funding from HCIA R2 to expand the MSHOP to 39 non-UMHS sites.

The MSHOP taught participants scheduled for an abdominal surgery healthy habits, giving them a standardized kit that contained a pedometer, spirometer, and materials encouraging them to walk more, perform breathing exercises, stop smoking, eat foods that promote health, and reduce their stress. To encourage participants to engage with the program without placing much burden on the practices, the awardee designed an automated tracking system that followed up with participants—daily via telephone or text, or weekly via email—to remind and encourage them to record their numbers of steps and breaths in a day. The automated tracking system also enabled participants to log their progress online or by telephone or text.

Based on existing literature about the benefits of prehabilitation as well as findings from a UMHS study on the MSHOP that took place before the cooperative agreement, the University of Michigan believed the MSHOP program services would lead to fewer surgical complications and reduce the length of inpatient hospital stays after surgery, both of which would lower costs (Table I.1).

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The MSHOP was designed to assess participants' risks for poor outcomes after an abdominal surgery and, if medically appropriate, to engage them in a prehabilitation program that would potentially improve their surgical outcomes.
Major innovation	The MSHOP's key innovation was to educate pre-surgical patients at high risk for poor surgical outcomes to positively change their health behaviors by using tools and materials they received in a prehabilitation kit. In addition, the automated tracking system designed by the awardee was an innovative way to follow up with MSHOP participants to remind and encourage them to use the tools in the kit and to log their progress online or by telephone or text.
Program components	Quality improvement and process redesign, health information technology, and patient and family engagement
Target population	Individuals at participating practices who met the following criteria: (1) were scheduled for a major abdominal surgery, (2) were scored as high risk for poor surgical outcomes, and (3) had at least one week between the MSHOP enrollment and the surgery date. The three-year enrollment projection was 2,500 total participants.

[3] Santa Mina, D., H. Clarke, P. Ritvo, Y. W. Leung, A. G. Matthew, J. Katz, J. Trachtenberg, S. M. H. Alibhai. "Effect of Total-Body Prehabilitation on Postoperative Outcomes: A Systematic Review and Meta-Analysis." *Physiotherapy,* vol. 100, no. 3, 2014, pp. 196–207. doi:10.1016/j.physio.2013.08.008.

Moran, J., E. Guinan, P. McCormick, J. Larkin, D. Mockler, J. Hussey, J. Moriarty, and F. Wilson. "The Ability of Prehabilitation to Influence Postoperative Outcome After Intra-Abdominal Operation: A Systematic Review and Meta-Analysis." *Surgery,* vol. 160, no. 5, November 2016, pp. 1189–1201. doi:10.1016/j.surg.2016.05.014.

Table I.1 *(continued)*

Program characteristic	Description
Theory of change/ theory of action	The awardee hypothesized that the MSHOP's education, tools, and consistent reminders would motivate participants to engage in the prehabilitation activities of walking, completing breathing exercises, reducing or quitting smoking, improving nutrition, and reducing stress during the time between their enrollment in the program and their surgery. The University of Michigan believed that engaging participants in prehabilitation activities would lead to fewer surgical complications and would reduce the length of inpatient hospital stays after surgery, both of which would lower costs.
Payment model	Fee-for-service
Award amount	$6,389,850
Effective launch date	September 15, 2014
Program setting	Surgical practices
Market area	A mix of urban, suburban, and rural
Market location	Michigan
Target outcomes	The goals (based on results from a test of this program at UMHS) were to reduce: • The length of inpatient hospital stays by 2.3 days per case • The payments to hospitals for inpatient cost of care by $2,561 per case • Surgical complications by 10 percent

MSHOP = Michigan Surgical and Health Optimization Program; UMHS = University of Michigan Health System.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was partly successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, the awardee enrolled 2,226 participants—89 percent of its enrollment target—by the end of the initial cooperative agreement. Second, the awardee delivered its two program services as intended. Practice staff or the MSHOP coordinating center successfully (1) delivered MSHOP kits to participants and (2) monitored participant progress through automated or live communication. Third, the awardee adequately staffed the MSHOP coordinating center. Fourth, although the awardee was able to engage practice staff in the MSHOP, it had trouble securing effective physician champions and engaging surgeons throughout the cooperative agreement. Lastly, the awardee engaged the majority of MSHOP participants in the program and, on average, engaged them for longer than the minimum of one week.

Impact evaluation. Due to too few treatment beneficiaries, we do not anticipate being able to conduct a rigorous impact analysis for University of Michigan. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. After the cooperative agreement ended, the University of Michigan transferred ownership of the MSHOP to the Michigan Surgical Quality Collaborative (MSQC).[4]

[4] For more information, see http://www.msqc.org/.

Participating practices will use care management codes (G-codes) to bill for MSHOP enrollment through Blue Cross Blue Shield of Michigan (BCBSM) and its adjunct plans such as Medicare Advantage.

Sustainability. Under the purview of the MSQC, some of the MSHOP features will be changed post-award based on lessons learned by the University of Michigan during the cooperative agreement. Some changes will eliminate less effective program features; others are designed to make the program more comprehensive and attractive to participants and providers. In addition, the program was being scaled to all general surgery patients and to pre-operative planning for pain management at the University of Michigan.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July to October of 2016 with a sample of 44 potential respondents and achieved a response rate of 76 percent. The clinician survey was fielded from March to June of 2017 with a sample of 146 potential respondents and achieved a response rate of 70 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partly successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

The University of Michigan used three levels of identification, recruitment, and enrollment to secure hospitals, their affiliated practices, and patients to participate in the MSHOP. First, the awardee leveraged its existing relationship with the MSQC to get the collaborative's permission to recruit only MSQC hospitals into the MSHOP. Once the University of Michigan received this permission, it presented information about the MSHOP to hospitals and invited them to participate in the program.

Next, for hospitals that expressed interest, the awardee met with hospital leaders to secure data use agreements (DUAs) and institutional review board (IRB) approvals, and met with clinicians and staff at hospital-affiliated practices to get MSHOP buy-in and conduct training on the MSHOP technology and how to incorporate the MSHOP procedures into practice workflows.

Lastly, practices that agreed to participate and completed training were expected to start identifying, recruiting, and enrolling eligible patients into the MSHOP during surgical consults. Specifically, the MSHOP targeted adults at participating surgical practices who (1) were scheduled for a major abdominal surgery, (2) were at high risk for poor surgical outcomes, and (3) had at least one week between MSHOP enrollment and surgery. To identify high risk patients, the University of Michigan encouraged non-UMHS surgeons and their staffs to use a tool on a mobile device or laptop during surgical consults, or at points of referral, to assess patients' risks for poor surgical outcomes.[5] Non-UMHS surgeons and their staffs used either the tool or their clinical impressions to identify high-risk patients and invite them to participate in the MSHOP, whereas all UMHS surgeons used their clinical impressions to identify high risk patients to recruit into the program.

If surgeons and their staffs did not use the risk assessment tool during surgical consults or referral points, they or staff at the MSHOP coordinating center used it after patients were enrolled in the MSHOP to ensure that only those program participants at high risk for poor surgical outcomes were counted in the HCIA R2 cooperative agreement. While the awardee enrolled both participants who were at high and low risk for poor surgical outcomes, low risk MSHOP participants (that is, HCIA R2-ineligible MSHOP participants) did not receive program services funded by HCIA R2.

b. Evidence of enrollment effectiveness

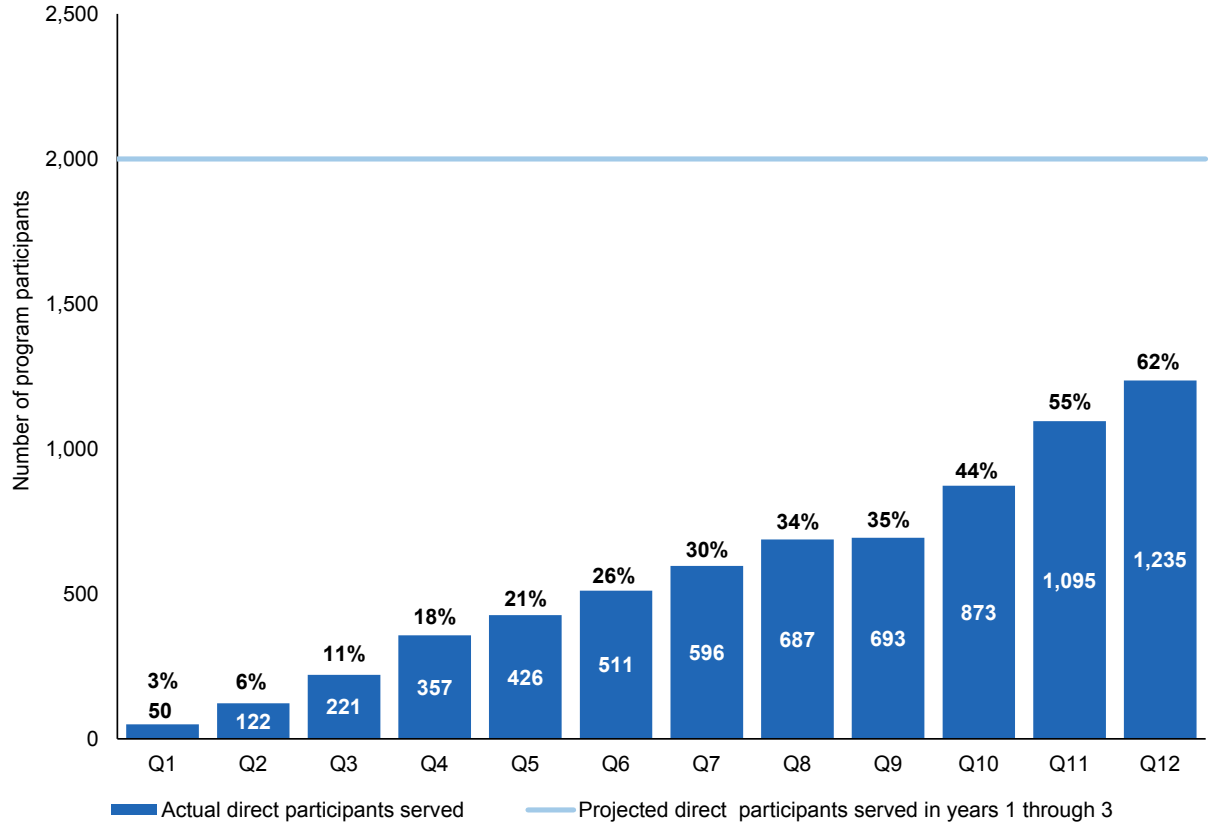
Overall, the University of Michigan reported that it enrolled 2,226 direct and indirect participants from September 2014 through July 2017, which represents 89 percent of its final three-year projection (Figures II.1 and II.2).[6] The awardee originally established a target for enrollment of 12,500 direct and indirect participants by the end of the three-year cooperative agreement, but lowered its target from the original goal twice, first to 5,243 in March 2016 and then to 2,500 in August 2016. The awardee reported that its original target was based on MSQC data on hospitals' surgical volumes, which were not analyzed at the level of detail necessary to make accurate enrollment projections. This contributed to the awardee's difficulty meeting its projections throughout the cooperative agreement.[7] At the end of the first program year, the awardee had met less than 5 percent of its then-current enrollment target of 12,500 direct and indirect participants, and at the end of the second program year, it had met less than 22 percent of its then-current enrollment projection of 5,243 direct and indirect participants.

[5] The risk assessment tool used a predictive algorithm that was created using MSQC surgical data. Each participant received a risk score based on the outcomes of patients who had had the same surgery. Those who scored at or above 50 were considered high risk for poor surgical outcomes.

[6] Although the cooperative agreement ended in August 2017, the University of Michigan stopped enrolling patients in the MSHOP in July 2017. Participating practices could choose to enroll patients into the program after July, but activities related to these patients were not funded by HCIA R2, and these participants were not counted toward the enrollment goal reported in this narrative.

[7] For more information on how the University of Michigan made its original projection, see the second annual report at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

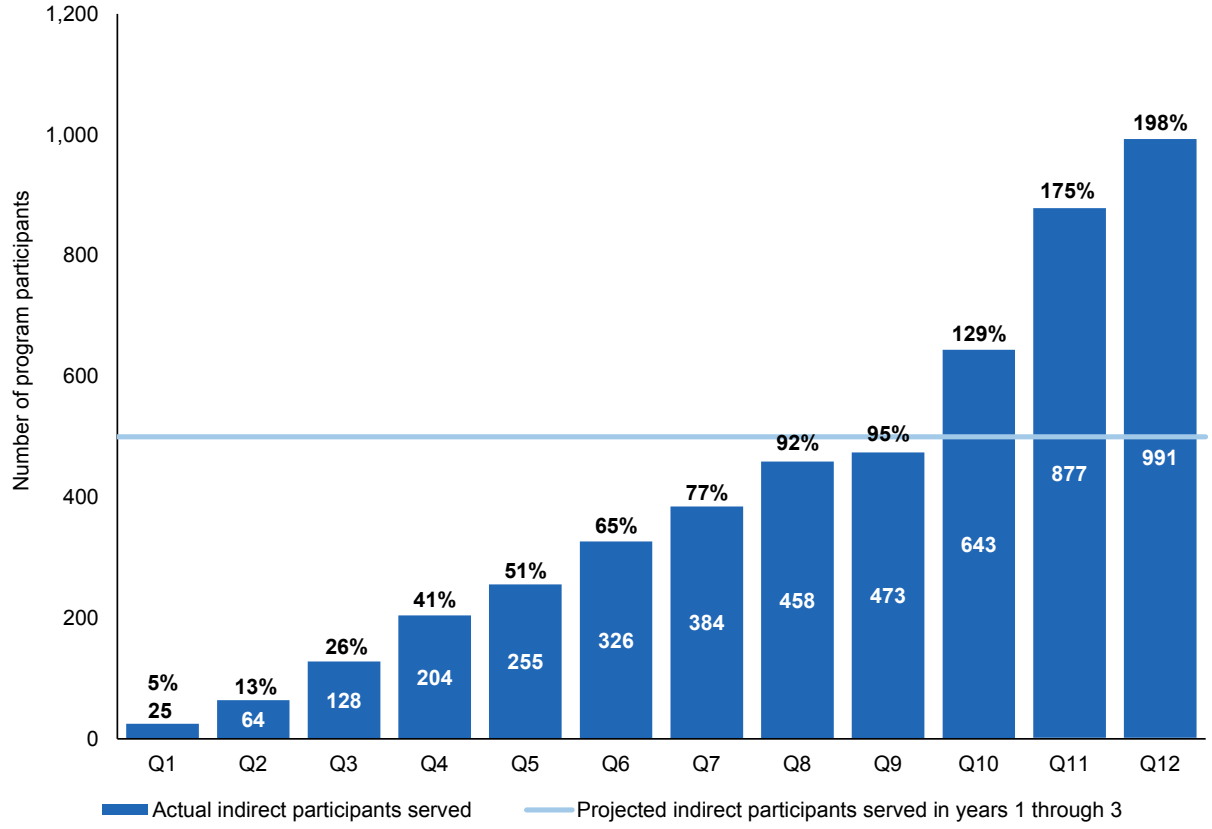
Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee lowered its direct participant enrollment projection several times during the three-year cooperative agreement, from 10,000 at the beginning of the cooperative agreement to 4,194 in March 2016 to 2,000 in August 2016.

Figure II.2. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers. The awardee lowered its indirect participant enrollment projection several times during the three-year cooperative agreement, from 2,500 at the beginning of the cooperative agreement to 1,049 in March 2016 to 500 in August 2016.

c. Facilitators and barriers associated with enrollment effectiveness

The University of Michigan's progress in meeting its three-year enrollment goal was influenced by two facilitators and four challenges.

The existing relationship between the University of Michigan and the MSQC gave the awardee easy access to a group of non-UMHS hospitals and affiliated practices that were already engaged in quality improvement activities and were likely to be interested in participating in the MSHOP because it aligned with the goals of their work. In addition, UMHS implemented the MSHOP in 2011, giving it time to address enrollment challenges and refine mitigation strategies for MSHOP implementation well before the cooperative agreement. In fact, from program inception to March 31, 2017, 63 percent of all HCIA R2-eligible MSHOP participants counted in the cooperative agreement were enrolled at UMHS as opposed to non-UMHS sites, indicating

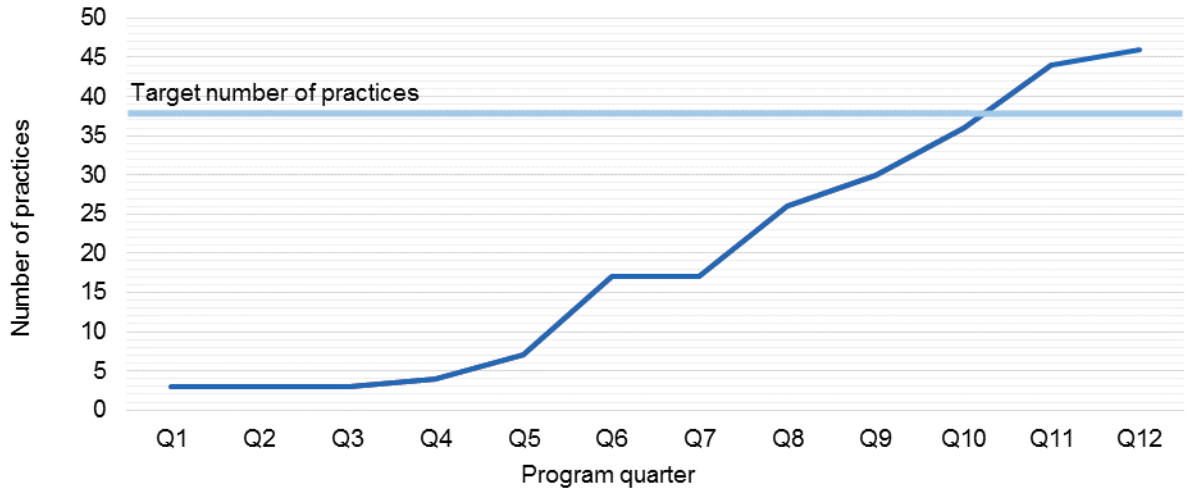
that the extra time the health system had to implement the program was a facilitator to enrollment.

Although the University of Michigan successfully enrolled participants into the MSHOP at UMHS, it experienced four challenges that affected enrollment at non-UMHS practices.

First, lengthy processes to secure DUAs and IRB approvals for hospitals led to a slow enrollment pace at affiliated practices and, in turn, a slow enrollment pace for patients. The awardee initially planned to enroll 40 hospitals by the end of the cooperative agreement. In the first program year, after learning it would take more time than anticipated to enroll hospitals because of the DUA and IRB processes, the University of Michigan changed the site enrollment goal to 40 surgical practices instead of having a goal for the number of hospitals to enroll. This allowed the awardee to focus on enrolling practices in hospitals that already had agreements and approvals in place.

Although making this change made it easier for the University of Michigan to recruit and enroll practices, the awardee still endured lengthy DUA and IRB processes at some hospitals throughout the cooperative agreement. As shown in Figure II.3, the University of Michigan ultimately exceeded its revised goal of recruiting and engaging 40 surgical practices by the end of the cooperative agreement, but did not reach the goal until close to the end of the third program year. This meant that late-joining practices had less time to enroll patients into the program.

Figure II.3. Number of participating practices, by program quarter



Sources: Awardee narratives, Q1–Q12.

Notes: The University of Michigan considered a practice fully recruited to the MSHOP if it had administrative approval to participate in the program; had completed the administrative agreements and business documents; had been trained on the use of the MSHOP technology, and was ready to use the MSHOP technology from a technical standpoint.

The University of Michigan Health System (UMHS) began implementing the MSHOP before the cooperative agreement. UMHS is counted as one "practice" because the health system consists of surgical lines or teams rather than individual practices.

Figure II.3 *(continued)*

Near the end of the first program year, the University of Michigan formally revised its site recruitment goal to recruit 40 surgical practices instead of 40 hospitals. This allowed the awardee to focus on recruiting practices within hospitals that already had the necessary DUA and IRB approvals in place.

The University of Michigan recruited 48 practices, but two dropped out due to their low volumes of potentially eligible patients and lack of continued interest in the MSHOP. These two practices are not included in the figure.

Second, a delay in the release of the risk assessment tool, as well as surgeons' resistance to using it, slowed the pace of patient enrollment. Non-UMHS surgeons were initially required to use the tool to enroll patients, but it was not ready for use until May 2015—seven and a half months after the program launch—due to administrative issues. Moreover, once the risk assessment tool was ready to use, many non-UMHS surgeons preferred to use their clinical impressions, instead of the tool, to identify and enroll patients.

In fact, in the clinician survey many physician champions said that using the MSHOP web-based tool, which included the risk assessment tool, was a challenge to program implementation. In response, the University of Michigan allowed 14 of the 46 participating practices to opt out of using the tool to enroll patients.[8] This count includes UMHS, which stopped using the risk assessment tool before the cooperative agreement, also because of resistance from surgeons. However, even if surgeons and their staffs did not use the risk assessment tool during surgical consults or referral points, they or staff at the MSHOP coordinating center used it after participants were enrolled in the MSHOP to ensure that only those program participants at high risk for poor surgical outcomes were counted in the HCIA R2 cooperative agreement.

Third, many surgeons and practice staff at non-UMHS practices did not remember to recruit and enroll eligible patients into the MSHOP. This was because the MSHOP was not the standard of care for all inpatient surgical patients, and surgeons and practice staff saw eligible patients infrequently. In particular, this was an issue for surgeons and staff at practices affiliated with small hospitals with few eligible surgical patients.

To address this, in the second program year the awardee revised its eligibility criteria to include more types of abdominal surgeries, specifically expanding its original eligibility criteria from surgeries represented by 73 current procedural terminology (CPT) codes to encompass surgeries for 292 CPT codes.[9] It also allowed 34 of the 46 sites to enroll into the MSHOP patients scheduled for surgeries not on the CMS-approved list of abdominal surgeries. This count includes UMHS, which established the MSHOP as the standard of care for all inpatient surgical patients before the cooperative agreement. (Participants undergoing surgeries not on the CMS-approved list did not receive program services funded by HCIA R2 dollars and were not counted

[8] Because many surgeons did not want to use the risk assessment tool to enroll patients, the University of Michigan placed less focus on giving practices morphomics-based risk assessment reports, which were intended to help practices better understand each patient's risk score for a particular surgery. For more information, see http://www.med.umich.edu/surgery/morphomics/.

[9] The University of Michigan, with CMS approval, revised the eligibility criteria three times in the second program year of the award. The first revision expanded the eligibility criteria to include 690 CPT codes. Out of concern that the criteria had become too broad, the second and third revisions narrowed the criteria: first to 319 CPT codes and then to 292 CPT codes. For more information on the University of Michigan's changes to the list of eligible surgeries, see the second annual report at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

toward the enrollment goal.) In addition, the University of Michigan invited 26 non-UMHS practices that were struggling to enroll patients to participate in a centralized enrollment process: practice staff gave the MSHOP coordinating center lists of patients they wanted enrolled in the MSHOP, and MSHOP staff then called the patients to formally enroll them in the program. The count includes UMHS, which established the centralized process before the cooperative agreement.

Lastly, surgeons at smaller sites often performed surgeries less than two weeks after surgical consults, which eliminated some patients who would have otherwise met the MSHOP enrollment criteria. The University of Michigan learned that smaller hospitals have the ability and a financial incentive to schedule elective surgeries to be performed soon after surgical consult, whereas UMHS often schedules surgeries to be performed four to six months from the time of the surgical consult. The University of Michigan addressed this challenge by reducing the required gap between enrollment in the program and surgery from two weeks to one week. This change led some clinicians and program staff to question whether one week was enough time for the prehabilitation activities to have an effect on surgical outcomes. However, the awardee's data show that the average length of time a non-UMHS patient participated in the MSHOP was about three weeks, longer than the original minimum of two weeks (see Section II.B.2.b).

2. Delivery of program services

a. Description of and changes to service delivery model

The University of Michigan's program included two services. First, **practice staff gave participants MSHOP kits**—which contained educational material on the benefits of exercising, performing breathing exercises, stopping smoking, eating healthfully, and reducing stress before surgery—along with tools to help participants engage in prehabilitation activities (a pedometer, a spirometer, a summary of foods that promote health, and best practices on smoking cessation and stress reduction). The kits were distributed in person by practice staff to each patient who verbally expressed, during surgical consult or at the point of referral, that he or she wanted to participate in the MSHOP. The awardee allowed practices to choose whether they wanted to discuss the contents of the kits with participants or distribute the kits without much explanation. As a result, some participants received more in-person guidance on the program, though the awardee anticipated and allowed for this variation in its program design.

Second, **practice staff or the MSHOP coordinating center followed up with participants** after participants received the MSHOP kit. Practice staff or the MSHOP coordinating center relied on an automated tracking system that reminded participants daily or weekly to submit their data by telephone, text, or online. This included how many daily steps they logged on their pedometer and how many breathing exercises were recorded via spirometer each day. If participants did not respond to the automated communication within one week, practice staff or the MSHOP coordinating center followed up with them over the telephone or in person at medical appointments.

b. Evidence of service delivery effectiveness

The University of Michigan was partly successful at implementing its HCIA R2 program in four domains. First, the awardee delivered its two services to participating patients largely as intended throughout all three years of the cooperative agreement. Second, the awardee

adequately staffed the MSHOP coordinating center. Third, although the awardee was able to engage practice staff in the MSHOP, it had trouble securing effective physician champions and engaging surgeons throughout the cooperative agreement. Lastly, the awardee engaged the majority of MSHOP participants in the program and, on average, engaged participants for longer than the minimum of one week. We provide details on implementation effectiveness below.

Delivery of intervention services. The University of Michigan delivered both the MSHOP kits and follow-up with participants mostly as planned. First, in interviews throughout the cooperative agreement, the University of Michigan, surgeons, and practice staff reported that the MSHOP kits were given to all patients who expressed interest in participating in the program during surgical consult or at the point of referral. It is important to note that, although the University of Michigan designed the kits to be standardized across sites, it allowed 5 of the 46 practices to modify the kit. For example, practices that were already distributing a spirometer may have substituted their own brand for the MSHOP spirometer or added supplemental educational materials to the kit. The awardee ensured that customized kits contained the same information as standardized MSHOP kits by requiring practices to receive approval from the MSHOP coordinating center before making any changes. Interview respondents in the third program year said these changes did not hinder the awardee's ability to deliver the educational materials and tools necessary for patients to understand and participate in the MSHOP.

Second, the University of Michigan reported in the third program year that follow-up with participants was delivered mostly as planned. Throughout the cooperative agreement, the University of Michigan used the automated system to follow up with participants on their prehabilitation activities and collect their numbers of steps taken via pedometer and breathing exercises conducted with the spirometer. The awardee initially planned for practices to both track the engagement of their participants via the web-based patient tracker and to follow up with any participants who were not recording their activities. However, some practices were burdened by these tasks. In response, the University of Michigan offered 45 of the 46 practices the option of centralized follow-up, in which staff at the MSHOP coordinating center contacted non-responding participants on behalf of practices.

Staffing and training. The University of Michigan hired people to fill 16.2 full-time equivalent (FTE) positions over the course of the cooperative agreement, which is 69 percent of its target. Despite not meeting its staffing target, the awardee reported that it generally had enough staff to implement and manage the MSHOP throughout the cooperative agreement. Beginning in the first program year, the awardee changed its staffing structure and strategy in response to the shortage of FTEs and began cross-training staff members at the MSHOP coordinating center so they were no longer limited to distinct job roles and could help with the most pressing MSHOP tasks. For example, there were up to four people on the team who enrolled patients, up to three people who conducted in-person trainings for practices, and two people who were able to conduct all major tasks when necessary.[10] As a result

> "We've always had to function as a team where everybody does a little bit of everything, and people are cross-training."
>
> —MSHOP leader

[10] Given the small size of the team, the University of Michigan did not set goals for training its staff. Staff were trained by working with their managers and other colleagues.

of these changes, the University of Michigan reported that it only needed 16 FTEs to carry out the MSHOP tasks during the practice recruitment stage and 10 FTEs to manage MSHOP-related work at the practices after that.

Recruitment and engagement of providers. The University of Michigan sought to engage three types of providers: (1) physicians to serve as champions in leading MSHOP implementation, (2) surgeons to recruit and enroll patients, and (3) practice staff to carry out other MSHOP activities such as delivering the MSHOP kits to participants.[11] The awardee reported, in interviews throughout the cooperative agreement, that it experienced difficulty securing effective physician champions and engaging surgeons, but that it was able to effectively engage practice staff in the MSHOP.

Even though the University of Michigan offered physician champions the potential to earn annual incentive payments of $1,500 along with the opportunity for their practices to earn $2000 to $4,000, the awardee reported that these payments were not enough of an incentive for clinicians to get deeply involved with using the MSHOP. This was confirmed in results for the survey of clinicians. Few physician champions and clinicians said they were motivated to participate in the MSHOP because of the incentives that were offered.

In terms of the program's administrative requirements, however, there were some discrepancies between the awardee's perceptions that it was difficult to engage the clinicians and the views of the clinicians themselves. Only a few physician champions or clinicians reported in the clinician survey that the MSHOP administrative requirements, such as meetings and documentation, were unreasonable. In addition, 90 percent of clinician respondents said the MSHOP did not increase their feelings of burnout at work. These data indicate that clinicians, particularly physician champions, may have lacked a clear understanding of what the University of Michigan expected of them.

Engagement of program participants. The University of Michigan successfully engaged program participants in two ways. First, the main metric the awardee used to monitor participant engagement with the program was submission of the numbers of steps and breaths taken, using either the automated technology or live communication. The University of Michigan reported that it engaged about two-thirds (67 percent) of both HCIA R2-eligible and -ineligible MSHOP participants between program inception and March 31, 2017 (Table II.2).

Because the University of Michigan recruited and enrolled patients at low risk for poor surgical outcomes, as well as patients scheduled for surgeries that were not on the CMS-approved list of 292 CPT codes, the awardee did not distinguish between HCIA R2-eligible and -ineligible groups in its engagement of program participants.

Second, the University of Michigan's data on average length of patient engagement with the MSHOP indicate that participants engaged in the program not only longer than the required minimum of one week, but also longer than the original minimum of two weeks. The University of Michigan reported that the length of time HCIA R2-eligible participants engaged with the

[11] The University of Michigan engaged some physicians as program champions although the MSHOP was centered on surgeons and their surgical staff. For example, one anesthesiologist participated in the MSHOP as a champion.

MSHOP averaged 32 days for the period ranging from program inception to March 31, 2017. (The average for participants at non-UMHS practices was nearly 23 days, and the average for participants at UMHS was a little more than 37 days.)

Table II.2. Levels of MSHOP participant engagement for both HCIA R2-eligible and -ineligible participants, from program inception to March 31, 2017 (n = 2,298)

	High engagement[a]	Low engagement[b]	Unenrolled[c]	Did not engage[d]	Had surgery in less than one week[e]
Percentage (number) of participants	59% (1,356)	8% (180)	17% (384)	10% (236)	6% (142)

Sources: Awardee-reported data as of May 2017.

Notes: The denominator of 2,298 includes MSHOP participants who had scheduled surgery dates before or on March 31, 2017, including participants not counted in the cooperative agreement (that is, low-risk participants and/or those who were scheduled for abdominal surgeries not on the CMS-approved list of 292 CPT codes).

It is important to note that the level of MSHOP patient engagement could not have reached 100 percent. A few MSHOP participants had to drop out of the program due to factors outside their control, such as changes in their diagnosis and emergencies that required them to have their surgeries sooner than planned.

[a]Participants submitted data for their numbers of steps and breaths into the tracking system at least three times per week for at least half of the weeks they were enrolled in the MSHOP.

[b]Participants submitted data into the tracking system, but not often enough to meet the "high engagement" definition.

[c]Participants dropped out of the MSHOP because, for example, their surgeries were canceled or postponed or because they chose to drop out.

[d]Participants never submitted their data into the tracking system.

[e]Participants ended up having surgery less than seven days from the day of enrollment.

c. Facilitators and barriers associated with service delivery effectiveness

The University of Michigan's progress in achieving its service delivery goals over the three-year cooperative agreement was influenced by three key factors—specifically, one facilitator and two challenges.

First, the University of Michigan gave practices flexibility by working with them one-on-one to determine MSHOP implementation protocols that were suitable for them to incorporate into their workflows. For example, the University of Michigan allowed practices to decide if the enrollment process would work best in the surgeon's office, the practice's affiliated prehabilitation clinic, or another type of central clinic—such as a pre-admission testing clinic or preoperative clinic. Another example of the awardee's flexibility is that, as noted, the University of Michigan invited struggling non-UMHS practices to participate in centralized enrollment and/or centralized follow-up processes in which practice staff gave the MSHOP coordinating center lists of patients they wanted enrolled or followed up on, and MSHOP staff then called the patients on behalf of the practices. The awardee noted that the centralized processes led these practices to view the MSHOP as simple and flexible. In addition, two practices said the centralized enrollment process helped engage participants. The practices believed that having the MSHOP coordinating center follow up with patients one or two days after they were introduced to the MSHOP during their surgical consults gave the patients time to process information about

their surgeries, and ensured that they started the program when they were emotionally prepared to.

Second, the University of Michigan found it challenging to effectively engage physician champions and surgeons in the MSHOP throughout the cooperative agreement. The University of Michigan attributed the difficulty to the intensive nature of these clinicians' surgical work, which limited their ability to focus on the MSHOP. For example, the University of Michigan reported that it shortened its training time on the MSHOP for physicians and surgeons to 15 minutes, as opposed to the 60-minute training for staff, because physicians and surgeons were unwilling to commit more time to training. Also, as discussed in Section II.B.1.c, many surgeons did not remember to recruit and enroll eligible patients into the program. The awardee introduced new activities in the third program year to secure and increase clinician engagement, launching both a social media campaign using Facebook and Twitter and a blog to share best practices on prehabilitation with physicians and surgeons and to reinforce their commitment to the MSHOP. It also hosted an event that brought physician champions together to discuss the facilitators of and challenges to MSHOP implementation.

Although the University of Michigan reported that the in-person meeting was effective in motivating some clinicians to increase their engagement with the MSHOP, the social media campaign was less successful. The awardee concluded that many physicians and surgeons do not use Facebook, Twitter, and other types of social media to learn about quality improvement programs. Data from the survey of clinicians suggest that the MSHOP did not greatly impact practice team interactions, the work environment, or the job satisfaction of program staff. Though the MSHOP's purpose was to improve patient care (many clinicians who responded to the survey said they believe the MSHOP did improve care) these data indicate that a lack of direct benefit to the practice may have been a factor in the awardee's difficulty securing engagement from physician champions and surgeons. It is possible that, if clinicians believed that practices would directly benefit from the MSHOP in the short term instead of indirectly through their patients' surgical outcomes in the long term, they may have been more motivated to engage with the MSHOP.

A second challenge to effective service delivery was the University of Michigan's difficulty in staffing the MSHOP coordinating center. It was not able to quickly fill open positions due to the university's hiring regulations.[12] In addition to this hiring problem, one of the University of Michigan's two clinical program liaisons resigned at the end of the second program year.

The University of Michigan identified two lessons learned about staffing the MSHOP. First, an MSHOP leader stated that the MSHOP coordinator center could have had more success working with practices to implement the MSHOP if it had more staff with clinical backgrounds, who could have worked more effectively with practices to implement the MSHOP into clinical workflows.

Second, the awardee said that providing FTE support to practices would have reduced implementation burden on practice staff and reduced the need to rely on volunteer efforts from

[12] For more information on the University of Michigan's challenges with hiring, see the second annual report at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

clinicians and program staff. In addition, the awardee stated that providing FTE support to practices would have allowed it to more closely manage site-specific implementation and potentially mitigate challenges more quickly. One survey staff respondent also recommended that the practice have one to two staff members who are formally dedicated to managing MSHOP activities.

C. Assessment of perceived program effects on the delivery of care and outcomes

Most of the clinicians and program staff who responded to surveys thought the MSHOP had positive impacts on the delivery of care and health outcomes. However, clinicians and staff lacked consensus on how the MSHOP led to positive impacts on care delivery and health outcomes. In addition, a few interview respondents believed the MSHOP was only somewhat effective in meeting its goals.

> "If you look at [how] … we made impact, I think [it is in] connecting the patient to be part of the care—the delivery of their care. Patients can directly impact the outcomes of their care. [That] is really the driving factor."
>
> —MSHOP leader

Most clinicians and staff who responded to the surveys, as well as those who were interviewed, thought the MSHOP had positive impacts. As shown in Table II.3, 9 in 10 clinicians and staff who responded to the survey thought the program had positive impacts on aspects of care delivery and health outcomes, the most common being better preparation of participants for surgery. Also, more than three-quarters of the interview and survey respondents (76 percent of staff and 81 percent of clinicians) noted positive impacts on patient satisfaction with care. In addition, many of the clinicians and staff (79 percent and 66 percent, respectively, in the surveys) believed the MSHOP was making a difference in meeting critical needs in their area or community. Overall, most clinicians (89 percent) surveyed said they would recommend the MSHOP to a close friend or relative who needed care or services, believing that prehabilitation enhanced both patient satisfaction and surgical outcomes as it empowered and engaged participants. Lastly, most clinicians (87 percent) said they think the MSHOP should be implemented in other clinical settings or workplaces.

Table II.3. Clinician and staff perceptions of MSHOP's impacts

Survey item	Clinician respondents	Staff respondents
Percentage of clinicians and staff indicating that the MSHOP has had a positive impact on the following:		
Their ability to better prepare patients for surgery	90	90
Quality of care and services they provided to patients	84	81
Patient satisfaction	81	76
Patient quality of life	77	87
Their ability to provide care or services that were responsive to patient preferences, needs, and values	77	76

Source: HCIA R2 Clinician Survey, 2017, and HCIA R2 Staff Survey, 2016.

Note: Not all clinicians and staff responded about these items because the question was not relevant or applicable to certain clinicians and staff. Therefore, the results are not based on the full sample of clinicians and staff.

Although clinicians and staff believed the MSHOP had positive impacts on the delivery of care, they lacked consensus on what the MSHOP's most critical service to participants was and how it led to positive impacts on care delivery and health outcomes. For example, when they were asked about the means through which the awardee was effective in achieving its goals, responses varied as to whether it was through improved patient education or better patient treatment plans and outcomes.

A few interview respondents believed the MSHOP was only somewhat effective in meeting its goals. For example, one thought the program would reduce surgical complications, but not to a statistically significant degree. Another said the goal of reducing the average length of inpatient hospital stays was achievable in the aggregate because large facilities like UMHS participated in the MSHOP, but that it was not achievable for smaller facilities with lower volumes of eligible patients.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

The findings from our implementation evaluation have four important implications for the impact analyses, and will weaken our ability to conduct a rigorous difference-in-differences impact evaluation. First, UMHS implemented the MSHOP in 2011, three years before the cooperative agreement, so we cannot identify a pre-period for the intervention. Second, we cannot replicate the risk assessment tool used to identify patients at high risk for poor surgical outcomes, and are therefore unable to identify a credible comparison group. Third, it is important to note that MSHOP participants whose surgeons used the risk assessment tool may have had systematically different levels of risk than participants whose surgeons did not use the tool. Fourth, given that the University of Michigan expanded the types of abdominal surgeries to be included in the eligibility criteria in the second program year, less major surgical procedures may have been added to the eligibility criteria, thereby diluting the overall treatment effect.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of the University of Michigan's MSHOP program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect, (2) whether claims can identify the primary expected effects, and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: University of Michigan

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	796[a]
Projected number of Medicaid FFS enrollees with 6 months of program exposure by February 28, 2018	0[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,164
Likelihood of all-cause hospitalizations	763
MDE sample size requirement to detect 20% effect	
Total expenditures	291
Likelihood of all-cause hospitalizations	191
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Questionable, many participating surgeons were providing intervention services prior to HCIA R2 cooperative agreement
Claims sufficient to identify treatment and comparable comparison group?	No, low rate of identification of treatment group with claims data with significant dilution of treatment effect within an intent-to-treat framework
Likelihood of solid comparison group	Serious concern. We may be not able to identify a strong comparison group because enrollment was heavily based on clinical judgment
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None

Table III.1 *(continued)*

Evaluability domain	Response
Primary reason for no rigorous evaluation	Too few treatment beneficiaries for primary outcomes. The pre-existing pilot study makes it difficult or impossible to define a pre-period. Also, study eligibility criteria cannot be replicated in claims data.
Survey data for treatment group that will be analyzed	Clinician and staff surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We do not plan to conduct a rigorous impact estimate of the MSHOP. The number of projected enrollees is slightly greater than the number needed to detect an effect of 20 percent on Medicare expenditure, which suggests that undertaking a full analysis might be worthwhile. There is, however, an additional obstacle that further complicates the analysis. The HCIA R2 program represented an expansion of an existing pilot program carried out by the University of Michigan Health System. The majority of beneficiaries who appear in the finder file and undergo an eligible surgical procedure were treated at practices in the earlier pilot program. The pre-intervention period for these practices would therefore need to be several years earlier than for practices in the MSHOP itself. Moreover, many surgeons chose not to use the risk assessment tool to identify eligible beneficiaries, which was a key component of the new program. Instead, they continued to exercise clinical judgment when deciding which patients would benefit from the intervention. Using clinical judgment to enroll patients makes it difficult to identify a strong comparison group. Finally, we have not been able to identify a matched surgical Medicare claim for many beneficiaries. The awardee explained that this may be a result of the cancellation or rescheduling of their surgeries, which is often not indicated on the finder files.

B. Characteristics of Medicare and Medicaid participants at baseline

The baseline characteristics presented in this report are based on the Medicare FFS beneficiaries who were enrolled in the MSHOP (that is, screened and found to be at elevated risk for postsurgical complications) and whose name was sent to us by the awardee. The event that sets prehabilitation services in motion is an outpatient consultation between a patient and a participating surgeon at least one week before a scheduled, qualifying abdominal operation. In July 2016, the University of Michigan received approval from CMS to use an expanded list of eligible abdominal procedures for selecting candidates for the intervention. During the consultation, the surgeon's office staff invites the patient to enroll in the prehabilitation program. If the patient agrees to join, he or she is enrolled during the consultation. The prehabilitation intervention usually lasts for one to two weeks after a patient enrolls until the operation is performed. If the surgery is rescheduled, then the intervention period may last several weeks longer. The awardee expects that most qualifying surgeries will be performed on an inpatient basis.

As of the end of May 2016, the awardee had enrolled 1,022 participants.[13] According to the information on the finder file, 41 percent (419 individuals) were Medicare-only beneficiaries, 8 percent (78 individuals) were Medicaid-only beneficiaries, and 1 percent (10 individuals) were dually eligible for Medicare and Medicaid. According to the finder file, the remaining half of participants (515 individuals) had other sources of health care coverage, or they were uninsured.

We excluded 278 beneficiaries who received prehabilitation services for procedures not included in the expanded list of eligible surgical procedures approved by CMS in July 2016. At the end of the second program year, the awardee proposed to expand the number of eligible abdominal surgical procedures in an effort to increase enrollment and began to enroll patients scheduled for the proposed procedures before receiving approval from CMS. Ultimately, CMS approved most, but not all, of the newly proposed procedures. As a result, the awardee had to retroactively identify on the finder file those patients who were already enrolled even though their procedure was not approved and mark them as ineligible for services funded through HCIA R2. We also excluded another 24 enrollees because their participation in the intervention was suspended. According to the awardee, enrollees can be suspended for many reasons. They can withdraw voluntarily, they can change surgeons, they may die before their operation is performed, or their surgery can be cancelled or postponed until an unspecified date. Of the remaining 720 enrollees on the finder file, we were able to find 504 in the Medicare enrollment database (more than the number of awardees reported as Medicare beneficiaries on the finder file).

In presenting the baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they were enrolled in the awardee's program on or before May 31, 2016, the last time we updated baseline characteristics until the final study population will be available for analysis of baseline characteristics. The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. After we excluded beneficiaries who did not meet the above criteria, 222 participants were included in the analysis of baseline characteristics for this report.[14] Most of those who were excluded were dropped because they were enrolled in Medicare managed care or they were enrolled in MSHOP after May 31, 2016.

[13] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

[14] We were able to find a professional claim from a participating surgeon for an eligible procedure after the anticipated surgery date for 140 of these 222 matched beneficiaries. Missing claims could occur for a variety of

In terms of demographic characteristics, the University of Michigan enrolled a fairly representative group of Medicare beneficiaries (Table III.2). Over half of the participants (55 percent) are age 65 to 74, and nearly one-quarter of them (23 percent) are older than 74. Slightly more than half of the participants (55 percent) are female. Most of the participants (89 percent) are white. Blacks represent the next largest racial group at 9 percent. The participants are predominantly Medicare-only beneficiaries (82 percent); only 18 percent are eligible for both Medicare and Medicaid.

reasons. For example, (1) the operation had not yet been performed, or it was performed but the claim had not been submitted, (2) the surgery was billed under a different procedure code or under a different provider ID, or (3) the procedure was performed on an outpatient basis.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the University of Michigan's program through May 31, 2016

	All participants (N = 222)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	48	22
65 to 74	123	55
75 to 84	43	19
85 and older	8	4
Gender		
Female	122	55
Male	100	45
Race		
White	197	89
Black	19	9
American Indian, Alaska Native, Asian/Pacific Island American, or other	3	1
Hispanic	1	0.45
Original reason for Medicare eligibility		
Old age and survivor's insurance	148	67
Disability insurance benefits	72	32
End-stage renal disease (ESRD)[a]	2	0.9
Hospice[a]		
Medicare/Medicaid dual status, percent dual[b]	39	18
HCC score[c]		Statistic
Mean		2.32
25th percentile		1.09
Median		1.79
75th percentile		3.17

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which a beneficiary consented to participate in the program, which usually occurs at least one week before a scheduled surgery. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Despite the demographic similarity of MSHOP participants to Medicare beneficiaries in general, the health care utilization and spending of MSHOP participants is well above the U.S. average. A large minority of participants (32 percent) became eligible for Medicare because of a disability. Furthermore, the average hierarchical condition category (HCC) risk score for MSHOP participants is more than twice the average HCC score for Medicare FFS beneficiaries nationally. Table III.3 lists a common set of cost and utilization measures, including core measures from the Center for Medicare & Medicaid Innovation. The University of Michigan's primary goal for its target population is to reduce surgical complications by 10 percent. Because surgical complications often lead to longer inpatient stays after surgery, the awardee is also seeking to reduce the length of inpatient hospital stays by 2.3 days per case and, in turn, to lower payments for inpatient cost of care to hospitals by $2,561 per case. We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare expenditures, in total and by major types of services. The total average PBPM Medicare expenditure during the baseline year was $2,188—nearly three times the 2014 national average for Medicare FFS beneficiaries of $792.[15] The average PBPM Medicare expenditure in the baseline year ranged from $1,585 in the second quarter to $3,420 in the fourth quarter, reflecting an increase in the intensity of services in the quarter immediately before enrollment. Medicare expenditures for inpatient services ($904 PBPM) were the largest driver of total cost of care for participants, followed by expenditures for outpatient services ($563 PBPM) and physician services ($463 PBPM).

[15] Except for ambulatory observation bed stays, national cost and utilization data are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the University of Michigan's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	222	211	217	222	222
Average Medicare expenditures PBPM[a]					
Total	2,188 (195)	1,614 (313)	1,585 (274)	2,066 (265)	3,420 (337)
Acute inpatient	904 (127)	727 (198)	639 (181)	807 (187)	1,418 (264)
Inpatient other[b]	25 (15)	51 (50)	17 (16)	34 (33)	0 (0)
Outpatient[c]	563 (67)	251 (32)	344 (78)	626 (99)	1,001 (97)
Physician services	463 (31)	332 (40)	359 (60)	428 (46)	721 (52)
Home health	81 (12)	68 (19)	72 (21)	76 (19)	108 (22)
Skilled nursing facility	75 (21)	123 (50)	71 (41)	26 (17)	82 (40)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	77 (15)	63 (18)	83 (28)	72 (16)	90 (18)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	897 (106)	697 (141)	618 (129)	841 (140)	1,405 (206)
Outpatient ED visits	915 (129)	833 (196)	805 (166)	988 (218)	1,027 (138)
Observation stays	163 (33)	136 (57)	112 (45)	238 (64)	162 (53)
Primary care visits in any setting	8,829 (580)	8,505 (899)	7,542 (686)	8,340 (697)	10,847 (918)
Primary care visits in ambulatory settings	6,399 (350)	6,084 (551)	6,008 (467)	6,218 (461)	7,243 (506)
Specialist visits in any setting	15,860 (928)	13,504 (1,469)	13,288 (1,094)	13,314 (1,104)	23,027 (1,204)
Specialist visits in ambulatory settings	12,133 (736)	9,765 (1,010)	10,705 (894)	10,114 (853)	17,694 (862)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	45 (3)	13 (2)	12 (2)	17 (3)	25 (3)
Percentage with an outpatient ED visit[e]	42 (3)	13 (2)	15 (2)	16 (2)	22 (3)
Percentage with an observation stay[f]	14 (2)	3 (1)	3 (1)	6 (2)	4 (1)
Percentage with a 30-day readmission among all discharges	24 (3)	23 (7)	30 (9)	12 (5)	28 (5)
Percentage of participants with a readmission among all participants	11 (2)	3 (1)	3 (1)	2 (1)	5 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. The enrollment date is defined as the date on which a beneficiary consented to participate in the program, which usually occurs at least one week before a scheduled surgery. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Baseline year expenditures and health care utilization figures include eligible Medicare FFS beneficiaries who received prehabilitation services for procedures not included in the list of procedures approved by CMS in July.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

The MSHOP participants' average use of expensive Medicare services was high before the surgical admission that precipitated enrollment. For instance, the annual inpatient hospitalization rate of 897 per 1,000 beneficiaries was over three times the national annual rate of 274 admissions per 1,000 beneficiaries in 2014. The annual rate of 915 emergency department (ED) visits that did not lead to a hospitalization per 1,000 participants was more than twice the 2013 national Medicare FFS rate of 445 per 1,000 beneficiaries. Furthermore, the annual observation stay rate of 163 per 1,000 participants was nearly three times the observation stay

rate of 56 per 1,000 beneficiaries for all Medicare FFS beneficiaries in 2013.[16] Forty-five percent of the MSHOP participants had at least one hospitalization during the year before enrollment; 42 percent also had an ambulatory ED visit at least once. The likelihood of a 30-day readmission among participants was also high (24 percent per discharge) compared with the 2014 national average for Medicare FFS beneficiaries (18 percent per discharge).

The higher rates of acute care service use and 30-day readmissions might reflect participants' poor health status overall and their general susceptibility to medical complications. The preoperative intervention is designed to reduce complications from scheduled surgery only; it is unlikely to have a significant effect on long-term costs or use of health care services.

At baseline, the annual rate of specialty service use for program participants in any setting (15,860 per 1,000 Medicare FFS beneficiaries) was substantially higher than the rate of primary care visits in any setting (8,892 per 1,000 Medicare FFS beneficiaries). This may be appropriate given the substantial health care needs of this population and the high need for specialty care services before major surgery.

Over the course of the four baseline quarters leading to enrollment, we observed a generally upward trend in the average PBPM total payments and in the average PBPM payments for inpatient, outpatient, and physician services. We observed a similar trend in rates of hospitalizations, outpatient ED visits, observation stays, and specialty services. Thus, beneficiaries targeted for this intervention are high-cost patients who often use acute care during the year before enrollment; they are also extremely high-cost patients who often use acute care services in the quarter just before enrollment. This trend has important implications for selecting a comparable cohort of Medicare FFS beneficiaries for the comparison group.

[16] See MedPAC, "A Data Book: Health Care Spending and the Medicare Program," June 2015.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

After the cooperative agreement ended on August 31, 2017, the University of Michigan transferred ownership of the MSHOP to the MSQC. Because the MSHOP aligns with the collaborative's existing work in quality improvement, as well as the collaborative's role in facilitating recruitment and enrollment of hospitals and affiliated practices, the University of Michigan believes the MSHOP will be successfully sustained under MSQC leadership. Participating practices will bill for the MSHOP enrollment using G-codes through BCBSM and its adjunct plans, such as Medicare Advantage.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, as well as previous narratives and payment model reports submitted to the implementation and monitoring contractor. We did not interview the awardee about its payment model because it proposed a fee-for-service approach.

B. Description of the payment model

As noted, MSQC owns the MSHOP as of September 1, 2017. Practices will use G-codes to bill for the MSHOP enrollment through BCBSM and its affiliated plans, including its Medicare Advantage products. As of August 2017, the University of Michigan was not developing plans with other payers. BCBSM will prorate payment for the G-codes based on time spent and type of practitioner conducting enrollment. The University of Michigan estimates payments will average $90 per patient for about one hour of work. A licensed practitioner operating under a physician's supervision can conduct enrollment and bill for the service. The definition of a licensed practitioner for this purpose includes dietitians, social workers, or nurses; it excludes medical assistants. BCBSM-insured participants will not incur costs.

The University of Michigan initially developed and implemented a model that was based on participation in the program and on a physician incentive program that existed between the awardee and BCBSM before the cooperative agreement began. The awardee paid incentive awards to 40 physician champions and 39 surgical teams in the three years of the cooperative agreement. In this system, physicians and surgical practice staff earned points for completing program processes and then received a percentage of the available incentive payment based on the total points earned in a given performance year. The physician champion was eligible for an award of $1,500 each year, and the team participation award ranged from $2,000 to $4,000. In the final year of the award, however, after receiving feedback from staff who felt overburdened by the MSHOP tasks, the awardee started to cover enrollment and patient follow-up costs for practices and began developing the new model based on G-codes.

C. Status of the payment model

The University of Michigan and BCBSM planned to conduct a six-month pilot of the model for three sites to help them determine how to do the necessary documentation and teach sites how to bill using G-codes. The goal was to go live to all participating sites in early 2018, but the timeline has shifted due to a delay in the revenue cycle at UMHS.

D. Factors associated with the development of the payment model

The awardee reported several challenges in implementing the original payment model based on incentives for participation. First, because of low enrollment across the program, few practices achieved the maximum incentive payment. Second, payments were too low to incentivize surgeons and staff to implement the program. Criteria for determining incentive payments were based on MSQC hospitals' surgical volume of all elective inpatient procedures and did not exclude elective procedures, or urgent, emergent, and other surgeries conducted within a two-week time window (precluding prehabilitation participation).

Based on its experiences with the original model, the awardee pursued the option to support MSHOP under existing FFS payment using G-codes. Program leaders reviewed three-month and six-month retrospective data for patients referred to the Michigan Medicine Preoperative Clinic, and found that billable G-codes would provide a substantial revenue source, surpassing the costs of the work.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The University of Michigan was working on its sustainability plan for the MSHOP, which first involved transferring the program to the MSQC. The awardee planned to pilot and implement its payment model by early 2018. The payment model was limited to beneficiaries of BCBSM and its adjunct plans, like Medicare Advantage. Some of the MSHOP components will be changed post-award, and the program was being scaled to all general surgery patients and to pre-operative planning for pain management at the University of Michigan.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, the awardee had begun implementing strategies to sustain the MSHOP. The awardee had strong institutional support from UMHS to sustain the program at the site, where the program started before the HCIA R2 award (in 2011). Non-UMHS practices varied on whether they wanted to continue the program after the cooperative agreement. The University of Michigan was also talking to BCBSM about the possibility of continuing the incentive payments to both UMHS and non-UMHS practices beyond the award period. In addition, the awardee scaled the program to recruit and enroll patients with any inpatient abdominal surgery, regardless

of whether the surgery was included on the CMS-approved list of surgical procedures. The awardee also planned to have the web-based tool available for a fee, through prenovo, to other providers interested in replicating the program.[17]

C. Implementing the SSR plan: progress and changes

Sustainability. In Year 3, the awardee decided to transfer the program to the MSQC, building on its existing relationship with the collaborative. The awardee needed to create closure plans and was trying to minimize service disruptions as much as possible by sharing tools and best practices with the MSQC. As noted in Chapter IV, the MSQC planned to sustain enrollment in the program for beneficiaries of BCBSM and its adjunctive plans by using G-codes for billing. This reimbursement would help the surgical practice employ someone to enroll and educate participants about the MSHOP and conduct the follow-up. The awardee and BCBSM were planning a limited, six-month pilot of the payment model, with full implementation to take place on January 1, 2018, but the timeline was shifting because of a delay in the revenue cycle at UMHS.

In sustaining the MSHOP after the cooperative agreement, the MSQC expected to make several changes to the program based on lessons learned by the awardee during the cooperative agreement. Some changes were intended to end less effective program features; others were designed to make the program more comprehensive and attractive to participants and providers. The program was ending the requirement that participants be risk-stratified. The MSHOP web-based tool was to be removed from the program, though practices could choose to independently work with prenovo, the company that developed the tool, to purchase the technology if they wanted to. The MSQC also expected to remove the automated follow-up component of the program. Instead, surgical practice staff would call participants to collect information on their prehabilitation activities. Lastly, the MSQC would expand the options for prehabilitation activities in the MSHOP (for example, swimming would be a new exercise option).

Scalability. The MSQC was also focusing on making the MSHOP the standard of care for all inpatient general surgeries at surgical practices. It planned to expand the program's patient eligibility criteria to include all general inpatient surgeries. Although they thought almost every patient could benefit from prehabilitation, program staff learned that the program is not necessarily meaningful for all participants (for example, those who are already athletically inclined would probably see little benefit). The program also expected to start including pre-operative planning for pain management, integrating best practices for prescribing opioids. A longer-term goal was to scale the program to encompass patients with outpatient surgeries.

Replicability. The awardee reported that the program would be applicable to other settings, but that it had not been replicated. The awardee's original plan for the other health systems to purchase the web-based tool to help them implement the MSHOP or their own version of it had not come to fruition.

[17] For more information, please visit http://prenovo.com/.

D. Factors associated with progress toward implementing the SSR plan

Respondents reported that key facilitators to sustaining the program were the positive preliminary results on cost savings and participant outcomes, and participants' testimonials on the program's value to them. According to the awardee, participants appreciated the follow-up activities because they showed that surgeons and their surgical staff cared about their well-being.

With the MSQC assuming responsibility for the MSHOP, the University of Michigan expected it to take time for both MSQC staff and surgeons and their staffs to learn how to operate and bill for the program under the new payment model. As noted, the program would no longer offer the web-based tool for free and would not include automated technology to collect data from participants about their activities. One program leader believed this could keep participants from being engaged in the program. However, the same respondent noted that, if the MSHOP generated revenue after the cooperative agreement, program leaders would encourage sites to use their own activity trackers to collect patient data if they did not want to purchase the web-based tool from prenovo.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. However, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or the sustainability, scalability, or replicability plans. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: University of New Mexico, Health Sciences Center

April 18, 2018

Elizabeth Coombs (Mission Analytics Group, Inc.)
Eric Verhulst (Mission Analytics Group, Inc.)
Eric Lammers (Mathematica Policy Research)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Due to lack of access to neurosurgeons and neurologists in New Mexico, many patients with neuro-emergent conditions, such as mild traumatic brain injuries or strokes, were being

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] During its one-year no-cost extension, the University of New Mexico will continue to enroll hospitals, fund tele-health consultations, evaluate the impact of the program, and build its payment model.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

transferred unnecessarily from their local hospitals to the University of New Mexico—the state's only Level I trauma, tertiary care provider—at great time and expense ($25,000 to $50,000 per transfer). The University of New Mexico used funding from HCIA R2 to launch the Access to Critical Cerebral Emergency Support (ACCESS) program, supporting hospitals in effectively treating patients locally through tele-health, often avoiding these unnecessary and costly transfers. Under the ACCESS program, when a patient presents in a participating hospital's ED with a neuro-emergent condition, the ED physician uses the Net Medical Xpress Solutions (NMXS) tele-health platform to communicate with a neurologist or neurosurgeon. These specialists, contracted by either NMXS or the University of New Mexico, examine the patient, review imaging, and discuss treatment options through the technology's secure file transfer and video capabilities. Based on the consultation findings, the patient is either discharged, admitted, or transferred. Tele-health coordinators located at each participating hospital facilitate the consultation process and serve as the primary liaison between ACCESS and ED staff at participating hospitals.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The University of New Mexico (UNM) implemented the ACCESS program to facilitate tele-health consultations for patients who present at a participating hospital's ED with a neuro-emergent condition.
Major innovation	Tele-health is an innovative approach for providing specialty care to individuals in rural areas. Lack of access to specialists means that patients must often travel long distances to needed care. ACCESS tele-health fills this gap by making specialists available to patients and their providers through innovative technology, including video monitors, diagnostic equipment, and scan-sharing capabilities. ACCESS also provides innovation through tele-health coordinators who act as program advocates within hospitals and UNM clinical staff who build hospital capacity to treat neuro-emergent conditions through training.
Program components	Tele-health
Target population	• Adults and children who present in a participating ED with a neuro-emergent condition. • Adults may be insured through Medicare, Medicaid, or not covered by either insurance. Children may be insured through Medicaid, CHIP, or neither.
Theory of change/ theory of action	UNM hypothesized that tele-health consultations would decrease the time it took for the patient to receive a treatment recommended by a specialist, decrease unnecessary hospital transfers, improve physician confidence in treatment decisions, and improve patients' satisfaction because they were being treated closer to home. In turn, the access to specialty care provided locally through tele-health would result in better health outcomes and lower health care costs.
Payment model	New fee-for-service (FFS) payment and shared savings
Award amount	$15,042,466
Effective launch date[a]	May 4, 2015
Program setting	Hospital ED
Market area	Initially rural, then expanded to the entire state
Market location	NM
Target outcomes	Lower health care costs due to a reduction in unnecessary hospital transfers

[a]After the initial planning period, the awardee's program became operational as of this date.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the University of New Mexico was partly successful at implementing its program by the end of the initial three-year cooperative agreement. Although the program faced challenges with enrollment, participating hospitals reported the ease of the technology and important outcomes. The University of New Mexico only served 74 percent of its target population through the ACCESS program, with 1,450 tele-health consultations. Not as many hospitals joined the program as anticipated because of concerns they could not cover the costs of the consultations for privately insured or uninsured individuals or after the cooperative agreement ended. In addition, lengthy specialist credentialing processes and staff turnover in participating hospitals negatively affected the number of consultation performed. That said, due to the University of New Mexico's partnership with NMXS, which developed and managed the tele-health platform, hospitals reported the tele-health process was straightforward and the technology easy to use. Hospitals also reported that the program helped them provide patients with timelier access to specialty care, improving patient outcomes and reducing transfers, which can be costly to insurers and disruptive for patients and their families.

Impact evaluation. Due to too few treatment beneficiaries, we do not anticipate being able to conduct a rigorous impact analysis for University of New Mexico. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. The University of New Mexico did not develop a new payment model, but instead identified a payment approach for neurological and neurosurgery tele-health services under existing FFS payment. The awardee is in discussions with a Medicaid managed care organization (MCO), and is aiming to develop an agreement with them for these services.

Sustainability. The University of New Mexico made progress sustaining its program beyond the cooperative agreement using five strategies: (1) implementing sustainable health IT infrastructure, (2) improving and optimizing program operations, (3) increasing the number of monthly consults, (4) working on policy and legal issues that affect the program's long-term sustainability, and (5) leveraging tele-health services to increase revenue. The awardee also scaled the program to two new sites by the end of the third program year and anticipated signing a contract with a new health system during the no-cost extension period.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July to October 2016 and achieved a response rate of 53 percent. The clinician survey was fielded from March to June 2017 and achieved a response rate of 67.3 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the University of New Mexico was partly successful at implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

Patients who arrived at a participating hospital's ED with a neuro-emergent condition were offered the tele-health consultation. Although patients could opt out of the service, the vast majority of them and their families chose to consult with the specialist through the tele-health technology.[4] All patients regardless of payer type were offered the consultation, though the cooperative agreement only paid for individuals insured through Medicare and Medicaid. The number of patients who participated in ACCESS was directly related to the number and size of hospitals enrolled in the program. No changes were made to this enrollment process.

[4] The University of New Mexico does not track the opt-out rate, but staff and hospitals reported that almost no patients opted out.

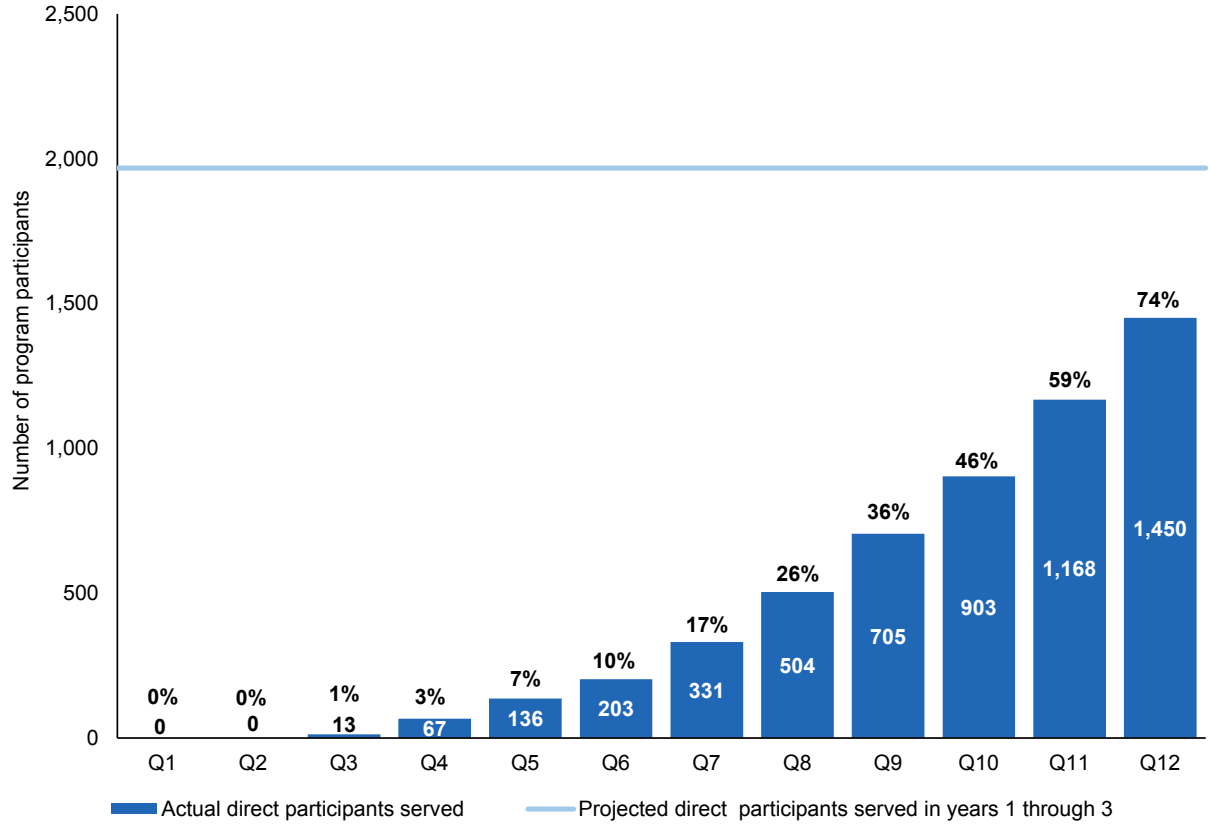
b. Evidence of enrollment effectiveness

The University of New Mexico faced challenges in achieving enrollment effectiveness. Overall, the University of New Mexico reported that the ACCESS program provided tele-health consultations to 1,450 patients enrolled in Medicare or Medicaid from May 2015 (when it launched its program) through August 2017—which represents 74 percent of its three-year projected participant total of 1,968 (Figure II.1). The University of New Mexico originally projected an enrollment target of 8,504 consultations, but decreased this target to 3,556 consultations in November 2015 and to 1,968 consultations in November 2016. Consultations varied from an average of 20 per month at larger hospitals to 1 or 2 per month at smaller hospitals. Over 600 individuals who were either uninsured or privately insured also received tele-health consultations.[5] Approximately 90 percent of all tele-health consultations were for neurology, whereas 10 percent focused on neurosurgery for conditions such as brain bleeds or swelling that may require surgical intervention.

As demonstrated by Figure II.1, the University of New Mexico steadily increased the number of consultations over the life of the cooperative agreement as more hospitals joined the program. However, ACCESS fell short of the target amount due to low hospital enrollment overall. The University of New Mexico aimed to initially enroll 30 hospitals in the program, but only succeeded in enrolling 12 hospitals. (Four additional hospitals may join the program in ACCESS's final no-cost extension year.)

[5] Uninsured and privately insured patients are reported as indirect in the figure.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The ACCESS program provided tele-health services to individuals who were privately insured or uninsured. These individuals are not presented in this figure. The University of New Mexico originally projected an enrollment target of 8,504 consultations, but decreased this target to 3,556 consultations in November 2015 and then to 1,968 consultations in November 2016.

c. Barriers and facilitators associated with enrollment effectiveness

The University of New Mexico's faced five barriers in making progress toward meeting its three-year enrollment goals. First, despite the University of New Mexico's ample hospital recruitment efforts, many hospitals were hesitant to join ACCESS for financial reasons. They were concerned that the hospital could not afford to take on the costs of the tele-health consultations for individuals not covered by the grant (uninsured and privately insured individuals). The University of New Mexico hopes more hospitals will adopt the ACCESS model once the financial benefits of treating patients locally are quantified and disseminated.

Second, both hospital and ACCESS program staff also reported that some hospitals were hesitant to join because of poor perceptions of governmental cooperative agreements. For one, hospitals were concerned with sustainability—that once the cooperative agreement was over,

they would be "left with nothing." Although University of New Mexico staff attempted to demonstrate that they were dedicated to the hospital partnership and the project regardless of funding source, hospitals were "gun shy" due to poor prior experiences. In addition, hospitals reported that burdensome paperwork and coordination with funders sometimes made it easier to "go with private industry."

Third, hospital plans to adopt alternative tele-health platforms also hindered hospital enrollment. Instead of paying a fee for every consultation with ACCESS ($600 to $1,200 depending on type and length of consultation), these hospitals aimed to pay an annual subscription, which might be more cost-effective for larger hospitals with higher utilization.

Fourth, staff turnover within these rural New Mexican hospitals resulted in multiple recruitment attempts per hospital, delaying enrollment. As reported by a University of New Mexico program manager "there is so much turnover in these hospitals that by the time we got something worked out with the CEO, he was gone and we had to start all over."

Finally, even once hospital contracts were signed, consultations were slow to start because the credentialing of specialists could take months. The University of New Mexico had hoped delegated credentialing would save time, but it only shortened the process by approximately two weeks. The University of New Mexico offered to provide technical assistance on credentialing to hospitals through screen-sharing technology and on-site visits, but much of the credentialing burden still fell on hospitals, which were short-staffed and inexperienced with the process.

The University of New Mexico aimed to improve hospital recruitment and thus increase the utilization of ACCESS by hiring a staff person with contacts throughout New Mexico. Given that New Mexico is a small state where "networking and knowing people is important," this staff person helped gain the trust and thus the participation of multiple hospitals. In the third grant year, the University of New Mexico hired an additional person to increase program marketing.

In addition, recruitment was facilitated by some hospitals' preexisting use of the NMXS tele-health technology. Under the grant, these hospitals upgraded the technology, added the neurosurgery component, and received training and support in providing neurological care. These hospitals were already familiar with the process and were happy to receive the added benefits under ACCESS.

2. Delivery of program services

a. Description of and changes to service delivery model

Once a hospital finalized the contracting and credentialing process, NMXS installed the technology and trained the tele-health coordinator and ED clinical and administrative staff on the process of making a request and facilitating the consultation. When a patient presented in a participating hospital's ED with a neuro-emergent condition, staff used the tele-health technology to send relevant scans and connect to the consulting specialist via video monitor. The specialist then recommended a treatment plan, which included advice on discharging, admitting, or transferring the patient. NMXS and the University of New Mexico responded to early lessons learned by streamlining and expediting the consultation process. They simplified the consultation request process through an online form, helped hospitals categorize acuity to prioritize urgent

cases, and trained hospitals on the difference between neurology and neurosurgery to enable them to make the appropriate requests.

b. Evidence of service delivery effectiveness

Despite challenges with hospital enrollment that led to fewer consultations than anticipated, the tele-health technology worked as intended and ACCESS was implemented effectively. Hospital staff reported that the consultation process was straightforward and allowed them to connect their patients to needed specialty care. In addition, the University of New Mexico provided hospitals with educational opportunities on care for neuro-emergent conditions. Below, we describe service delivery effectiveness in relation to the delivery of intervention services, staffing and training, engagement of providers, and engagement of program participants.

Delivery of program services. Both University of New Mexico and hospital staff reported that the NMXS tele-health platform was easy to use and experienced few glitches. The University of New Mexico hospital liaison communicated with hospital staff and NMXS to identify and address any issues with the technology or process. Although the frequency and severity of these issues were not systematically reported, both hospital staff and the University of New Mexico hospital liaison described few implementation challenges during interviews. Internet connectivity seemed to be the largest barrier. Although most related issues have been resolved, some hospitals continued to struggle due to their rural location.

Interviewees also indicated that the quality of treatment plans provided by specialists was high. A neurosurgeon and neurologist at the University of New Mexico reviewed about a third of the consultations to ensure that the recommended treatment plan was aligned with the patient's condition and needs. Interviewees noted that the University of New Mexico had to raise questions with hospitals or specialists or follow up with patients to ensure they received the most appropriate care for only a "few cases."

University of New Mexico and hospital staff also reported the positive impact of ACCESS on the administration of tPA. Although tPA can be life-saving for patients with ischemic strokes, physicians are often hesitant to prescribe the medication because it can be life-threatening for patients with brain bleeds. According to hospital and University of New Mexico staff, ACCESS patients received tPA more often and sooner because of the tele-health consultations. University of New Mexico nurses reported that in rural America, about 2 percent of patients with a stroke receive tPA within the recommended time frame, compared to 18 to 20 percent of ACCESS patients. The fast administration of tPA not only can save lives and reduce the long-term effects of strokes, but also can help a hospital avoid lawsuits for not providing timely and appropriate care.

Staffing and training. The ACCESS program experienced little staff turnover, pointing to implementation effectiveness related to staffing and training. The University of New Mexico team consisted of program leadership/management staff, nurse trainers, and a hospital liaison. Although there was some change in program management staff, the turnover did not negatively impact program implementation. In fact, according to the University of New Mexico staff, new program management helped with hospital recruitment due to preexisting relationships in the state. The nurse trainers and hospital liaison, who were with ACCESS since program launch,

were part of the University of New Mexico's neurology/neurosurgery team and easily transitioned into the program without extensive formal training.

Tele-health coordinators experienced higher turnover and less engagement than the University of New Mexico staff. Their primary role was to serve as tele-health advocates within local hospitals, orienting new clinicians to the technology and reporting back issues to the hospital liaison. However, according to University of New Mexico staff, some were unresponsive, which required program management and the hospital liaison to become more involved in hospital monitoring efforts.

Engagement of providers and provider organizations. For implementation effectiveness, it was essential for ED clinicians to stay engaged in the program and regularly use the tele-health consultation platform. University of New Mexico staff tracked the number of consultations per month by hospital and intervened if utilization tapered off. Although there were fluctuations, overall, the number of consultations per month continued to steadily rise. For example, in May 2015, when the program was launched, there were 30 consultations per month. This number increased to 50 in May 2016 and to 90 in May 2017. Although this increase is a function of additional hospitals joining the program, it also demonstrates that the continuing hospitals found value in the program and came to rely on the consultations as a regular part of their care.

The University of New Mexico offered numerous training opportunities through the life of the cooperative agreement. The awardee held annual in-person conferences with all ACCESS hospitals, and nurse educators conducted in-person workshops and quarterly presentations with clinicians on treating individuals with neuro-emergent conditions. The University of New Mexico also encouraged ACCESS hospitals to attend "grand rounds" to learn about cases as examples of neuro treatment strategies. However, these trainings were less well attended by hospitals because they typically took place during busy clinic hours and were open to the larger neurology and neurosurgery community.

Despite these numerous training offerings, only slightly more than half of the 23 hospital staff members who completed a survey on ACCESS participation indicated that they attended a formal training on ACCESS (see Chapter IV for survey details). Of these, approximately 80 percent indicated that the training helped them learn new skills and improve job performance related to ACCESS. Sixty percent of those that did not attend a training thought that it would have been helpful to get better oriented with the program.

Interviews also indicated that ED clinicians were engaged in the program because they valued specialist support, especially related to tPA administration. One hospital reported that "ED doctors and the hospitalists take great comfort in having a specialist advise them and give that instruction 'yep, it's time for the medication.' There's a lot of comfort in that."

Engagement of program participants. After the tele-health consultation, the University of New Mexico followed up with patients to ensure they were receiving recommended care. According to the most recent quarter of data (Q11), 40 percent of patients attended a recommended 7-day follow-up appointment and 70 percent attended a recommended 30-day follow-up appointment. Improving the rates of these follow-up visits was not necessarily in the scope of the ACCESS intervention.

ACCESS also started conducting community outreach at hospitals. Most recently, University of New Mexico nurse educators conducted a seminar on stroke warning signs with 40 community members at a small hospital in southern New Mexico.

c. Barriers and facilitators associated with service delivery effectiveness

The effectiveness of service delivery was affected by two facilitators: (1) a successful partnership with NMXS and (2) a contract with additional specialists. In contrast, the University of New Mexico experienced two barriers to implementing the ACCESS program: (1) hospital turnover and (2) the small size of participating hospitals.

First, ACCESS benefited from a strong partnership with NMXS, which provided an intuitive and useful technology for staff to implement the intervention. NMXS staff also played a large role in training hospitals and addressing technology issues. One hospital reported that "whenever we've gone through the troubleshooting processes with [NMXS], it's always gone very smooth and there's always somebody willing to help."

Second, at the end of the grant's second year, the University of New Mexico increased program capacity through a partnership with Blue Sky Neurology in Colorado. In response to more hospitals participating in the program and the higher utilization of tele-health consultations, the University of New Mexico contracted with the specialty group to increase the number of neurologists providing consultations and decrease wait times.

Although ACCESS had the technological and specialist infrastructure in place, staff turnover and the small size of participating hospitals placed barriers on implementation effectiveness. First, new ED clinical staff needed training on the tele-health consultation process; ACCESS struggled with identifying new staff and ensuring they were aware of the technology. Staff turnover may be the reason why only half of the hospital staff surveyed had attended a formal ACCESS training. Although getting new staff up to speed on the program was the role of the tele-health coordinators, some "wore many hats" in these small, rural hospitals and were not able to dedicate the time needed to the project. This meant that when University of New Mexico staff saw utilization trailing off, they had to follow up with hospitals to ensure that new staff were aware of the technology and started using it.

Second, new technology often requires regular use so staff become comfortable with the process. Some participating hospitals were quite small, with just one or two consultations a month. Therefore, the time lag between consultations could mean that hospitals forgot about or deprioritized the process.

ACCESS addressed these two challenges through continual outreach, education, and monitoring. During the third program year, the University of New Mexico increased ACCESS's travel budget to enable the nurse education team to visit hospitals more often and provide education on neuro-care. These informational sessions not only built hospital capacity to treat patients locally, but also reminded them of the use and benefits of the technology. To improve hospital engagement, the University of New Mexico also hired a new staff member to monitor utilization and conduct outreach. This individual tracked the number of consultations and followed up with hospitals to see if they were facing challenges with the technology or needed additional training.

C. Assessment of perceived program effects on the delivery of care and outcomes

University of New Mexico and hospital staff indicated through surveys and interviews that the program was a cost-effective method to improve care delivery. Of the non-clinician hospital staff who completed the survey (21), the majority reported ACCESS had a positive impact on the quality (76 percent) and efficiency (81 percent) of care provided. Over 70 percent reported that the program was "worth the effort" and they would recommend it to a colleague. Although few individuals completed the clinician survey, those who did agreed that ACCESS had a positive impact on the quality of care they provided.

Hospital staff interviewed indicated they appreciated that the ACCESS program allowed them to avoid unnecessary transfers and treat patients locally. They also felt this generated revenue for the hospitals. Staff reported that prior to the tele-health program, the hospital would have transferred about half of the patients. Awardee data on transfers avoided were not yet available across the entire ACCESS program due to challenges collecting data on transfers made to out-of-state hospitals. Hospitals also reported that patients were happier being treated closer to home, near their families and where they have "preexisting physician relationships." Several interviewees reported cases where CT scans indicated the patient had no hope of survival. Instead of being transferred by helicopter to another hospital, they could "spend their final moments with their families and loved ones."

D. Implications of implementation findings for the design and interpretation of an impact analysis

Tele-health consultations were meant to provide patients quicker access to specialty care and reduce unnecessary transfers to other hospitals. However, the University of New Mexico was slow to enroll hospitals, which may cause challenges in estimating ACCESS's cost savings due to a reduction in unnecessary transports. By the end of the second program year, only eight hospitals had implemented the tele-health technology for at least a full year of the program. In addition, some hospitals already operated tele-health prior to the ACCESS program. Although they received more advanced technology and training under the cooperative agreement, the implementation of ACCESS may not necessarily have affected the number of unnecessary transports for these hospitals.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of University of New Mexico's ACCESS program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

University of New Mexico received a 12-month no-cost extension through August 30, 2018 and enrollment of new participants will end February 28, 2018. The projected sample sizes in Table III.1 are based upon actual enrollment as of August 31, 2017 projected through February 28, 2018. Due to processing lags in Medicaid data, we have not confirmed that the projected 408 Medicaid beneficiaries will meet program eligibility for inclusion in our impact evaluation.

Table III.1. Assessment of HCIA R2 awardee evaluability as of August 31, 2017: University of New Mexico

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure as of February 28, 2018	780[a]
Projected Medicaid population with 6 months of program exposure as of February 28, 2018	408[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	3,811
Likelihood of all-cause hospitalizations	1,558
MDE sample size requirement to detect 20% effect	
Total expenditures	953
Likelihood of all-cause hospitalizations	390
Participation/Selection bias of concern	Yes, concern about selection bias due to provider judgment/non-claims-based criteria used to identify treatment group.
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, an event or utilization/expenditures used to identify treatment group

Table III.1 *(continued)*

Evaluability domain	Response
Likelihood of solid comparison group	Some issues, but probably surmountable. Expect to select a comparison group but may be for a small percentage of all patients
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Too few treatment beneficiaries for primary outcomes
Survey data for treatment group that will be analyzed	Staff and clinician surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

Given our projections at this point, we are not likely to be able to conduct a rigorous impact analysis, but we will continue to explore the feasibility of this depending upon the final enrollment numbers. We do not have any awardee-specific data on implementation to report; however, we will report on the experiences of awardee staff, clinicians, and participants, based on our surveys.

B. Characteristics of Medicare and Medicaid participants at baseline

This section provides a summary of common and awardee-specific claims-based baseline characteristics for the treatment group, which we measured during the 12 months before each beneficiary's enrollment date. The treatment group consists of adult (age 18 and older) beneficiaries in Medicare fee-for-service (FFS), Medicaid FFS, and Medicaid managed care who received ED services with tele-health consultation at eight New Mexico hospitals participating in the ACCESS program. The treatment group is identified in lists of participants from the awardee.

The University of New Mexico began to enroll Medicare and Medicaid beneficiaries in the ACCESS program in May 2015. As of May 31, 2016, the awardee had enrolled 331 direct participants among all payers.[6] About 60 percent of the direct participants were Medicare FFS beneficiaries, while the rest were Medicaid beneficiaries.

In presenting the baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who had met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they must have been enrolled in the

[6] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

awardee's program on or before May 31, 2016. The calendar period covered by the baseline quarters is determined by the enrollment date for each participant, and therefore varies by participant. We also restricted the treatment group to those participants for whom we could identify a precipitating ED visit in claims data. We dropped 8 participants who did not have ED or inpatient claims within three days of their enrollment date. After we excluded patients who did not meet the above criteria, a total of 184 Medicare FFS participants were included in the analysis of baseline characteristics for this report.

The results of our analysis indicate that the University of New Mexico is recruiting a population that is diverse in terms of age and gender, but that it is predominantly white (88 percent of the Medicare FFS sample [Table III.2]). The population generally has significant health care needs and high Medicare expenditures (Table III.3). Eighteen percent of participants are younger than 65, whereas 22 percent are 85 or older. Twenty-eight percent were originally eligible for Medicare because of a disability, which is greater than the 24 percent in the Medicare FFS population nationwide who were originally eligible because of a disability. Five percent have end-stage renal disease (ESRD). Participants are also more likely to be female (58 percent). Thirty percent of participants are dually eligible for Medicare and Medicaid, which suggests that their socioeconomic needs are substantial—particularly, considering that 18 percent of the general Medicare FFS population is dually eligible. Participants are substantially less healthy and have a greater need for care than the general Medicare FFS population, as evidenced by the fact that the participants' average HCC risk score is more than 86 percent higher than that of the average Medicare FFS beneficiary.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the University of New Mexico's program through May 31, 2016

	All participants (N = 184)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	34	18
65 to 74	46	25
75 to 84	63	34
85 and older	41	22
Gender		
Female	106	58
Male	78	42
Race		
White	161	88
Black	2	1
American Indian, Alaska Native, Asian/Pacific Island American, or other	9	5
Hispanic	11	6
Original reason for Medicare eligibility		
Old age and survivor's insurance	127	69
Disability insurance benefits	52	28
ESRD[a]	5	3
Hospice[b]	0	0
Medicare/Medicaid dual status, percent dual[b]	56	30
HCC score[c]		Statistic
Mean		1.86
25th percentile		0.97
Median		1.48
75th percentile		2.42

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the beneficiary had an index visit for an emerging stroke or brain injury in an ED participating in the ACCESS program. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category.

ACCESS participants had high Medicare expenditures and high rates of service use in the year before enrollment. In Table III.3, we report baseline expenditure and utilization data for a common set of measures, including the four core measures from the Center for Medicare and Medicaid Innovation. We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments, in total and by major types of services. The total average PBPM Medicare payment during the baseline year was $1,530—substantially higher than the 2014 national average of $792.[7] Average PBPM Medicare payments among participants for the following services were the largest drivers of the total cost of care: acute inpatient services, $513; outpatient services (including ED visits), $273; and physician visits, $367.

ACCESS participants had high average use of expensive Medicare services before the index ED visit that precipitated enrollment in the program. The annual rate of acute care hospitalizations was 567 per 1,000 Medicare FFS beneficiaries in the treatment group during the baseline year—well above the national annual average of 276 per year per 1,000 Medicare beneficiaries in 2014. The annual rate of ED visits not leading to a hospitalization was 1,285 per 1,000 beneficiaries, compared with the 2014 national annual rate of 454 per 1,000 beneficiaries. The difference suggests that frequent ED users are more likely to receive services from the ACCESS program, which is not surprising given that the program setting is EDs. In the baseline year, the likelihood of a 30-day readmission (19 percent per discharge) was comparable to the 2014 national average (18 percent per discharge) for Medicare FFS beneficiaries. In addition, the annual rate of primary care visits in any setting (9,659 per 1,000 Medicare FFS beneficiaries) was lower than the rate of specialty service use in any setting (10,426 per 1,000 Medicare FFS beneficiaries).

From the period that extends from 7 to 12 months before enrollment (baseline quarters 1 and 2) to the period that extends from one to 6 months before enrollment (baseline quarters 3 and 4), we observed a substantial increase in average PBPM total payments (approximately, a 55 percent increase from the average of quarters 1 and 2 to the average of quarters 3 and 4). There was also an increase in average PBPM payments for acute inpatient services (up by 121 percent), physician services (up by 22 percent), home health services (up by 72 percent), and skilled nursing facility services (up by 90 percent). In addition, we observed upward trends in the last baseline quarter relative to the previous three in average PBPM payments for outpatient care. We observed similar patterns in the rates of hospitalizations, outpatient ED visits, observation stays, primary care visits, and specialist visits during the baseline year. These increases indicate that beneficiaries targeted by the ACCESS program were high-cost individuals who were heavy users of acute care services in the year before enrollment, especially in the two quarters that immediately preceded enrollment. This trend has important implications for selecting a comparable cohort of Medicare FFS beneficiaries for the comparison group.

[7] The national data here and in the next paragraph are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the University of New Mexico's program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	184	178	180	184	184
Average Medicare expenditures PBPM[a]					
Total	1,530 (257)	1,311 (204)	1,040 (131)	1,610 (491)	2,029 (327)
Acute inpatient	513 (200)	336 (109)	266 (73)	572 (389)	760 (209)
Inpatient other[b]	70 (25)	108 (64)	0 (0)	133 (69)	39 (34)
Outpatient[c]	273 (38)	255 (51)	267 (47)	252 (44)	317 (45)
Physician services	367 (43)	342 (44)	316 (40)	360 (79)	444 (69)
Home health	148 (27)	121 (32)	95 (25)	168 (37)	204 (41)
Skilled nursing facility	119 (31)	108 (49)	55 (38)	88 (55)	222 (73)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	40 (7)	41 (8)	40 (7)	37 (8)	43 (8)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	567 (191)	431 (109)	383 (98)	508 (216)	783 (184)
Outpatient ED visits	1,285 (195)	1,133 (206)	1,148 (174)	1,016 (256)	1,826 (303)
Observation stays	184 (36)	68 (39)	158 (58)	133 (61)	370 (86)
Primary care visits in any setting	9,659 (753)	8,000 (660)	8,100 (667)	8,569 (739)	13,826 (1,428)
Primary care visits in ambulatory settings	7,396 (440)	6,459 (509)	7,043 (556)	6,825 (523)	9,196 (713)
Specialist visits in any setting	10,426 (1,104)	9,678 (1,023)	10,283 (968)	9,674 (1,708)	12,022 (1,258)
Specialist visits in ambulatory settings	8,074 (587)	7,502 (716)	8,573 (752)	7,553 (715)	8,652 (729)

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	32 (3)	9 (2)	8 (2)	10 (2)	13 (2)
Percentage with an outpatient ED visit[e]	53 (4)	22 (3)	23 (3)	15 (3)	29 (3)
Percentage with an observation stay[f]	16 (3)	2 (1)	4 (1)	3 (1)	9 (2)
Percentage with a 30-day readmission among all discharges	19 (4)	24 (11)	10 (7)	11 (7)	28 (9)
Percentage of participants with a readmission among all participants	7 (2)	2 (1)	1 (1)	1 (1)	3 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

As part of ACCESS, the University of New Mexico did not develop a new payment model, but instead pursued options under existing FFS payment. The payment option covered neurological and neurosurgery tele-health services

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

For HCIA R2 participants, award funding covered the fees for the following services: (1) the consulting neurologist or neurosurgeon; (2) the tele-health platform vendor, NMXS; (3) call center to initiate and close a consultation; and (4) administrative and infrastructure support. For non-HCIA R2 participants who were privately insured, the awardee billed hospitals for neurology and neurosurgery consultations. Each hospital had the option of submitting claims to the non- HCIA R2 beneficiaries' insurance; however, most insurance companies would not reimburse a tele-health consultation. Most participating hospitals still engaged in the program (and therefore covered the tele-health consultation fee themselves) because they were satisfied with having patients retained at their facility who would otherwise have been transported to a higher level of care.

During the cooperative agreement, there were no tele-health codes available for billing, so the cost was reimbursed by either cooperative agreement funds or the hospitals. A fee schedule for these services also did not exist. The most common Diagnosis-Related Groups (DRGs) used for these services were strokes, concussions, and migraines.

After the cooperative agreement, the cost components will remain the same, but certain features will change. Changes include the program no longer covering the cost of Medicare HCIA R2 participants, the cost of neurology consults increasing from $600 to $850, and the cost of neurosurgical consults increasing from a minimum of $600 to $1,200. Part of this fee covers the cost of the NMXS tele-health platform developed and maintained by NMXS as well as equipment at the participating hospital. The remainder covers the reimbursement of the attending neurologist or neurosurgeon. The awardee worked with a health economist to identify the post-cooperative agreement cost. The awardee mentioned that hospitals will bill payers, hoping they will at some point reimburse hospitals for the consults. Currently, hospitals absorb the cost of the consult, hoping to recover those costs by avoiding a transfer and providing care for the participant at their own facility.

C. Status of the payment model

The awardee is in discussions with a Medicaid managed care organization (MCO), and is aiming to develop an agreement where the MCO will reimburse the tele-health consultation for

their members. The awardee also plans to meet with other MCOs in the market in hopes of having multiple contracts in place with payers by January 2018.

The awardee worked with the American Medical Association (AMA) to identify new or existing Current Procedural Terminology (CPT) codes that can be used for reimbursement from commercial payers and is planning to submit the AMA application for new CPT codes during the no-cost extension. During our interview, the awardee was also looking into whether they could use modifier (GT) codes once the cooperative agreement ends and they negotiate agreements with payers. To get a GT code, the awardee needs an established fee schedule; they were also working on developing this toward the end of the cooperative agreement.

As of the last program quarter, the awardee was in discussions with the New Mexico Human Services Department to provide support for ACCESS that would also impact the proposed payment option. One area for discussion is the possibility of including shared savings opportunities into the payment model. The department is also considering mandating that the state MCOs reimburse hospitals for ACCESS consultation conducted for state Medicaid patients.

D. Factors associated with the development of the payment model

The awardee identified the following key facilitators for developing the payment model:

- **Pilot study.** The awardee said that their pilot with a small-volume hospital, which occurred before the cooperative agreement, was critical to refining the service delivery and payment model.
- **Key personnel.** The awardee reported that it was crucial to have a broad mix of expertise, including those with an understanding of hospital procedures, payers, insurers, health care finances, health economics, IT, credentialing, contracting, federal and state law, as well as those with contacts in organizations and departments potentially influential in furthering the program. In addition, the collection of doctors, staff, and faculty provided the environment, expertise, and vision necessary for the incubation and development of the payment model.
- **Supporting data.** For the program to be able to forecast continuing viability and long-term sustainability, the awardee required data and the resulting analyses that provided evidence for participating in the program and payment model.

One major challenge in developing a payment model was that there was not a fair market value for tele-neurosurgery services in existence when the awardee began its program. Although stroke neurology was rather common, neurosurgery did not have examples either in practice or in the literature from which to draw an analysis. Therefore, the University of New Mexico worked closely with another participating community hospital in the state to conduct a fair market analysis. This involved learning how to create fair market valuations and identifying key stakeholders in the marketplace to review the analysis. The awardee used the analysis to set the reimbursement for tele-neurosurgery consultations.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The University of New Mexico made progress sustaining its program beyond the cooperative agreement using five strategies: (1) implementing sustainable health IT infrastructure, (2) improving and optimizing program operations, (3) increasing the number of monthly consults, (4) working on policy and legal issues that affect the program's long-term sustainability, and (5) leveraging tele-health services to increase revenue. The awardee also scaled the program to two new sites by the end of the third program year and anticipated signing a contract with a new health system during the no-cost extension period.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, the University of New Mexico had planned and made progress in sustaining its program after the cooperative agreement. The awardee identified a payment amount for program services, although neither private payers nor Medicaid had agreed to cover the fee. The awardee reported that it was waiting to approach payers again after analyzing program data to quantify cost savings. The awardee had also nearly finished developing a billing system for the program's telemedicine services and had plans to work with

the AMA to establish procedures codes for tele-health services. The awardee was also communicating its commitment to a long-term partnership with participating hospitals.

C. Implementing the SSR plan: progress and changes

Sustainability. As of the third program year, the University of New Mexico reported using five strategies to prepare the program for sustainment beyond the cooperative agreement: (1) implementing sustainable health IT infrastructure, (2) improving and optimizing program operations, (3) increasing the number of monthly consults, (4) working on policy and legal issues that affect the program's long-term sustainability, and (5) leveraging tele-health services to increase revenue. First, the awardee developed and implemented new health IT infrastructure that would allow "record keeping and billing for the ACCESS payment model" after the cooperative agreement ended. Second, the awardee continually improved and optimized its program during the cooperative agreement to help it be more efficient and therefore sustainable beyond the cooperative agreement period. For example, the University of New Mexico streamlined the consultation request process through an online form that allowed hospitals to categorize patient risk. The awardee also trained hospital staff on neurology versus neurosurgery consultations, so they connected with the correct specialist the first time. Third, the awardee signed new contracts with program partners to increase the number of consults. The awardee anticipated that these contracts would result in 200 consults per month during the no-cost extension period—which they believe is the program's break-even point and will sustainably cover the program's cost. During the cooperative agreement, the awardee reported an average of 130 consults per month. Fourth, the awardee continued working with external stakeholders and partners on policy and legal issues that affect the program's long-term sustainability. For example, the awardee met with multiple government officials and legislators to discuss support for its payment model, engagement of Medicaid MCOs, and legislation for telemedicine. The awardee also began working with a local law professor on policy issues, such as submitting proposals for new CPT codes, investigating questions about medical liability, determining gaps in health insurance and case law, and writing legislation to support telemedicine. Finally, staff at the University of New Mexico and implementing sites reported their intent to leverage tele-health services to increase revenue. Although tele-health services are not reimbursable, they help retain patients and, therefore, generate hospital revenue to offset the cost of the tele-health consultation.

Scalability. By the end of the third program year, the University of New Mexico was working on implementing the ACCESS service to two new implementing sites and anticipated signing a contract with a new health system during the no-cost extension period.

Replicability. The University of New Mexico did not report plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

The University of New Mexico reported that it had adequately used the cooperative agreement period to sustain the program by developing a working payment model, replacing program technology with sustainable health IT infrastructure, proving the program's value, and continually optimizing program operations. The challenges the awardee reported threaten long-term but not short-term sustainability. The main challenge, also described in the payment model chapter, is engaging payers and legislators to offer better reimbursement for tele-health program services.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension that allows it to continue to enroll new beneficiaries. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted above, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Ventura County Health Care Agency

April 18, 2018

Edward Kako (Mission Analytics Group, Inc.)
Eric Verhulst (Mission Analytics Group, Inc.)
Nick Theobald (Mission Analytics Group, Inc.)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background and purpose of the HCIA R2 initiative

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[2] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Ventura County Health Care, a public health agency based in Ventura, California, used funding from HCIA R2 to create the Chronic Obstructive Pulmonary Disease (COPD) Access to Community Health (CATCH) program through which the awardee sought to provide home-based care of patients in Ventura County diagnosed with COPD. The program's innovations

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

included providing care for the COPD patient in the community setting thereby reducing the frequency and severity of exacerbations and the resulting, costly hospitalizations. Additionally, CATCH established an incentive program that encouraged primary care providers to implement the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines for the care of patients with COPD. The awardee believed that these guidelines were not widely known among health care providers before CATCH was launched. In implementing the guidelines, primary care physicians (PCPs) were (1) trained to accurately diagnose COPD in one of four stages of disease severity and (2) given specific, evidence-based strategies for treatment that are aligned with each of the five stages of the disease. Participants in the CATCH program received care and case management from the program's registered nurses (RNs) and Registered Respiratory Therapists (RRTs); both coordinated care with the PCPs and specialists.

Ventura County Health Care identified potential participants in multiple ways: (1) during an ED visit or when they were discharged from a hospital, (2) by searching for COPD diagnosis codes in the electronic health records (EHRs) of participating clinics, and (3) through physician and partner referrals. Once beneficiaries were accepted into the CATCH program, they were assessed by an RN or an RRT who also administered a spirometry test to identify the COPD stage. CATCH staff began enrolling Medicare and Medicaid beneficiaries in January 2015. The awardee's goal was to enroll 2,500 beneficiaries during the three-year cooperative agreement. Its program outcomes included (1) reduced health care costs from reduced ED and PCP visits, (2) increased access to health care, (3) reduced COPD exacerbations, and (4) improved quality of life. The CATCH program included a payment model, called CATCHpay, that incentivized providers to train in and follow evidence based treatment guidelines for COPD. Providers who met the CATCHpay criteria (attending trainings and implementing the GOLD guidelines) were eligible for incentive payments. Ventura County Health Care made one major change during the cooperative agreement: it expanded the scope of CATCH to establish Pulmonary Rehabilitation program for eligible beneficiaries with COPD.

As part of the project, CATCH staff conducted community outreach events to screen and educate beneficiaries at-risk for COPD (i.e., smokers) in addition to individuals who already had COPD. Further details are included in Table I.1.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	Ventura County Health Care implemented the CATCH program to improve the quality of care for Ventura County residents with COPD by enhancing provider and patient awareness of the GOLD guidelines and expanding access to health care and resources for patients with COPD. The awardee expects these improvements to decrease the incidence of avoidable visits to EDs and visits to PCPs.
Major innovation	Although the GOLD guidelines are the accepted standard of care for patients with COPD, many health care providers are not aware of them. The goal of CATCH was to educate providers to follow the best practices available, using spirometry tests to determine COPD staging in addition to assessing symptoms to determine the level of care to provide COPD patients. The innovation also led to the creation of a pulmonary rehabilitation clinic in Ventura County (the CATCH clinic), which will be sustained as a new service after the cooperative agreement ends.

Table I.1 *(continued)*

Program characteristic	Description
Program components	• Provider training • Patient and family engagement
Target population	Medicare and Medicaid beneficiaries in Ventura County with COPD and those at risk of COPD as a result of exposure to smoke and second-hand smoke
Theory of change/ theory of action	The CATCH program is expected to lead to better COPD management by teaching the GOLD guidelines to staff in family clinics. The program is also expected to lead to improved outcomes in pulmonary function and quality of life. Finally, the CATCH program will reduce ED and specialty visits, resulting in cost savings to CMS.
Payment model	Provider incentive; Shared savings, bundled or episode payment
Award amount	$4,136,499
Effective launch date[a]	9/1/2014
Program setting	Patients' homes, family clinics
Market area	Urban and suburban
Market location	Ventura County, CA
Target outcomes	• Reduced health care costs from reduced ED and PCP visits • Increased access to health care • Reduced COPD exacerbations • Improved quality of life

[a]After the initial planning period, the awardee's program became operational as of this date.

COPD = chronic obstructive pulmonary disease; ED = emergency department; PCP = primary care physician.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that Ventura County Health Care was partially successful at meeting its enrollment goals. The awardee enrolled 2,162 participants, or 86 percent of its target enrollment goal, by the end of the three-year cooperative agreement. The awardee also successfully implemented most parts of its program, including provider education, provision of spirometry units, in-home consultations with participants, and opening a pulmonary rehabilitation clinic. There were only minor setbacks in some services. Discovered early on in the intervention, the setbacks were related to making antibiotics accessible to participants and licensing issues with a smoking cessation program. Starting in Year 2, the awardee began to engage providers in facilities outside the Ventura County Health Care Agency system. Throughout the program, CATCH staff engaged with participants at a minimum of once a month via text message and more often through phone calls, home visits, or CATCH clinic visits for participants who needed additional care.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of Ventura County Health Care's CATCH program, the analysis is still in progress and not included in this report. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. Ventura County Health Care's CATCHpay payment model combined a discounted bundled payment for treating COPD patients with an incentive payment for PCPs to encourage them to follow evidence-based COPD clinical guidelines. The awardee built a

bundled payment model for treatment, but CMS continued to reimburse for services by using its standard fee for service billing. The treatment guidelines have been successfully integrated into the primary care setting; however, the incentive payments ended with the grant funding.

Sustainability. Ventura County Health Care made progress on sustaining parts of the program funded by the cooperative agreement. The awardee finished integrating GOLD guidelines into the Cerner EHR system, which will remain in place beyond the cooperative agreement. In addition, the existing clinics will absorb program participants, although they may not receive the same home-based services. Finally, the awardee reported that it would sustain the pulmonary rehabilitation program through standard Medicare billing. The awardee did not take steps to replicate or scale its program in the third program year.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July to October of 2016 and achieved a response rate of 92 percent (12 of 13 people). The clinician survey was fielded from March to June of 2017 and achieved a response rate of 51 percent (86 of 168 people). We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a] Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that Ventura County Health Care successfully implemented its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

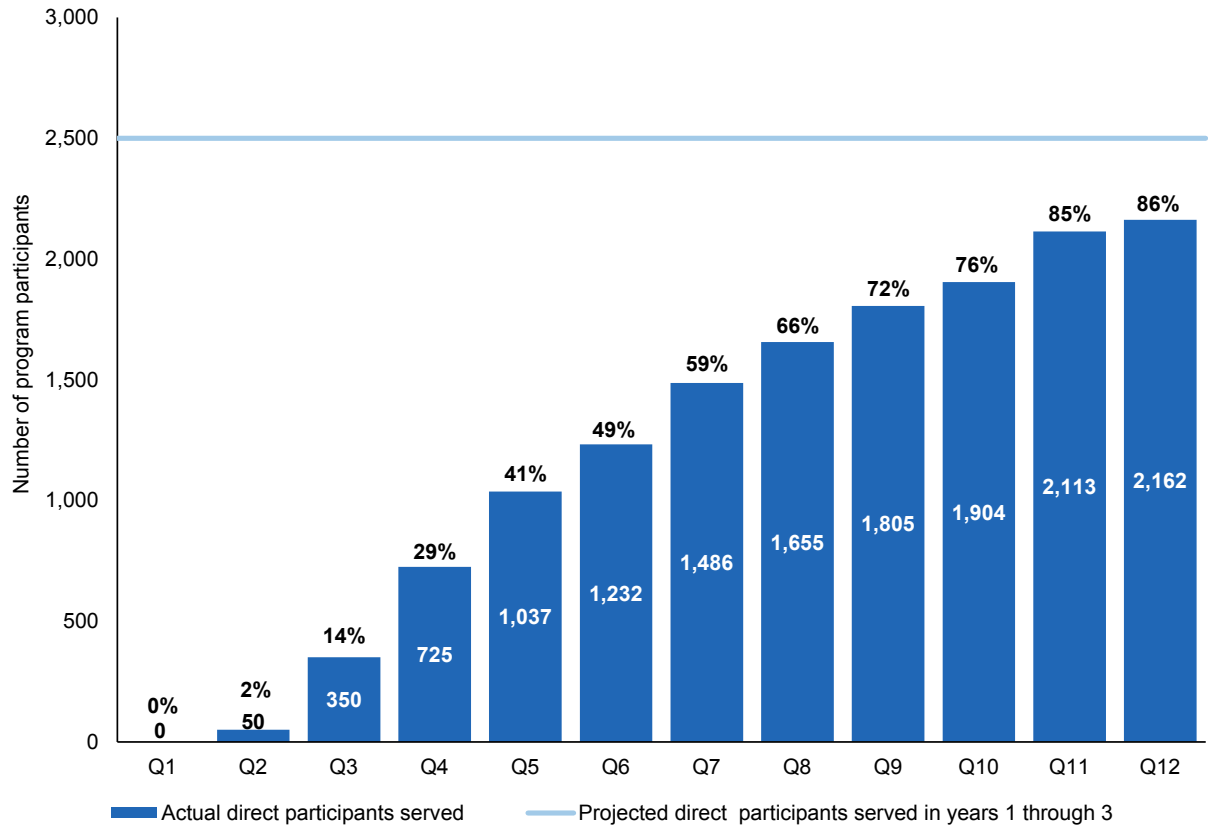
a. Description of enrollment strategies

The awardee used several strategies to enroll beneficiaries over the span of the program. In Year 1, CATCH staff focused on training providers within Ventura County Health Care Agency system to refer eligible beneficiaries by e-mail or phone. In Year 2, providers outside the county system began to refer beneficiaries. Also in Year 2, Ventura County Health Care updated the Cerner EHR system to integrate a pop-up referral option to CATCH as PCPs entered eligible diagnoses into the system. In addition, the awardee began providing nicotine replacement therapy (NRT) and evidence-based brief intervention (5 A's model) to people at risk of COPD (i.e., smokers).

b. Evidence of enrollment effectiveness

Ventura County Health Care began enrolling beneficiaries in January 2015. By January 2016, it had enrolled 1,130 beneficiaries; by January 2017, 1,867 beneficiaries had joined the program. The awardee reported that it enrolled 2,162 direct participants by the end of the program in August 2017. This represents about 86 percent of its final three-year projections (Figure II.1). The awardee ceased enrollment activities in July 2017 as it was approaching the end of the award period.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding.

c. Barriers and facilitators associated with enrollment effectiveness

The awardee listed one barrier to enrollment: bureaucratic hurdles to engaging providers outside the Ventura County Health Care Agency system. Beginning in Year 2, the awardee engaged partners outside the system to refer participants to the program. This was necessary to reach prospective participants outside the County system (and meet the enrollment target), but it was more complicated to share patient information outside the system. The awardee had to meet

requirements for each partner organization before it could begin enrolling their beneficiaries in the CATCH program.

Ventura County Health Care listed three facilitators of enrollment: (1) providing NRT to at-risk COPD patients, (2) providing spirometers and GOLD Guidelines training to providers, and (3) word of mouth and trust in the program.

First, Ventura County Health Care included the population eligible for services when it began providing NRT to at-risk COPD patients in Year 3. The awardee had to train staff to provide the NRT directly to participants. One of the CATCH RRTs explained that people responded better to tangible incentives: they were more likely to be interested in enrolling in CATCH if they were offered something directly.

> "I . . . noticed people want something in return, something concrete, and in the beginning, we were just giving out forms, like 'sign this form and we will provide you these services.' We were doing it like that in the beginning; I don't think that was sufficient. When we started offering screeners, like 'we're going to give you this test, just sign this for content, or we're going to give you this NRT to help you quit smoking,' that worked."
>
> —CATCH RRT

Second, the awardee noticed a similar pattern in providers earlier on. Much as NRT was an incentive for patients to enroll, offering spirometers to providers was an incentive for them to join the CATCH program, which led to more referrals and thereby increased enrollment. Spirometers can cost as much as $1,500 to $2,000. Covering the cost of the devices ensured that providers who were willing to provide pulmonary function tests (PFTs) could do so without having to supply their own funding.

> "It turns out the spirometers were more of an incentive to actually get providers to come on board and join us [than the CATCHpay bonus payments]."
>
> —CATCH RT

Finally, by Year 3, many stakeholders in Ventura County had heard of the CATCH program, which made it easier for Ventura County Health Care to (1) motivate additional providers to participate in the training on GOLD guidelines and (2) enroll beneficiaries into CATCH. Though some of the providers originally feared losing patients to CATCH, they became more trusting as they learned more about the program.

2. Delivery of program services

a. Description of service delivery model

The CATCH service delivery model included two components: provider training and patient and family engagement. Ventura County Health Care trained providers to integrate the GOLD guidelines into their practices when treating patients with COPD. Additionally, the GOLD Guidelines were built into the EHR to prepare for the switch from fee for service to the bundled payment model. The patient and family engagement component included a Patient-Centered Medical Home (PCMH) element: CATCH staff were sent to participants' homes to help them with their medication regimens, respiratory equipment, and any other issues related to COPD. After the initial home visit, CATCH maintained contact via text messaging with participants at least once a month. Participants with more severe COPD or who had additional needs could call CATCH staff at any time or could be seen in the CATCH clinic as necessary. As Ventura County Health Care began closing down the program, it connected participants with high needs to other case management services available in the county.

As mentioned at the beginning of this section, Ventura County Health Care began providing NRT to participants in Year 3, including those at-risk for COPD. At first, the awardee contacted the participant's PCP requesting the therapy on the participant's behalf. By the end of the cooperative agreement, some CATCH staff were trained to provide NRT directly to participants. One of the goals of providing free NRT was to more fully engage lower-stage COPD patients who had less need for other CATCH services by offering them a more tangible, relevant service. The awardee stated that providing free NRT also encouraged participants at all COPD risk stages to show up for their medical appointments.

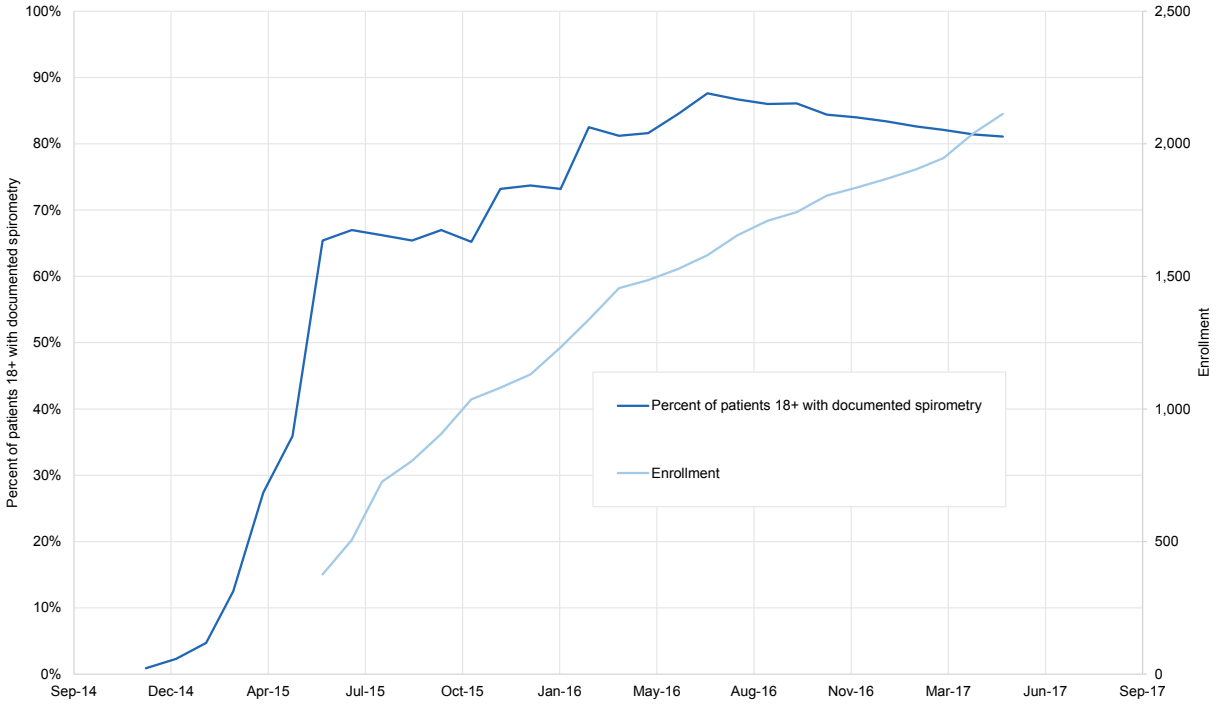
The awardee's theory of action was that increased education and training of both providers and participants would lead to reduced COPD-related exacerbations and thus fewer expensive outcomes, including ED visits and hospital stays.

b. Evidence of service delivery effectiveness

The CATCH program was successful in providing better care to patients with COPD in Ventura County. The awardee effectively delivered intervention service, recruited and trained staff, and engaged providers and participants. We based these conclusions on interviews with CATCH staff, which were supported by survey and self-reported measures data.

Delivery of intervention services. Ventura County Health Care successfully delivered intervention services. The awardee-reported data showed an improvement in many measures between baseline and the end of the program. Before the start of the program, less than five percent of participants had received a PFT. By the end of the program, the rate had risen to 82 percent. Figure II.2 shows the evolution of enrollment and the monthly rate at which PFT was administered among adults over the cooperative agreement. At first, there was a large increase in the rate at which adults received PFT. As more beneficiaries were enrolled, the rate stabilized because CATCH staff fell behind at the end of Year 1 in performing enrollment assessments (which includes the PFT) for new participants. The rate of spirometry assessment peaked at 88 percent in July 2016. One of the CATCH staff members interviewed in Year 3 mentioned that Ventura County Health Care had to re-order 70 spirometers because it ran out of units to give providers. It is possible that this temporary shortage led to a decrease in the rate of documented spirometry, since CATCH staff continued to enroll beneficiaries even though there were fewer resources to perform the PFTs. By May 2017, the PFT rate was 81 percent. By July 2017, there were only 10 spirometers left, which the awardee intended to distribute before the end of the program.

Figure II.2. Enrollment and percentage of patients older than 18 with documented spirometry by month as of August 31, 2017



Source: Enrollment and self-monitoring measurement data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Exact enrollment data were not available before June 2015.

CATCH staff also encouraged participants to join smoking cessation programs. As of the Quarter 12 report, the awardee reported a 17 percent increase in participation in smoking cessation programs from 37 percent before CATCH to 43 percent. Over the same period, the percentage of participants receiving inhaled bronchodilator therapy increased 85 percent, from 18 to 33 percent. Bronchodilator therapy is the crucial, first line of treatment for COPD.

Staffing and training. The awardee reported that the program was fully staffed beginning mid-way through year one of the cooperative agreement. Although there was a lot of work for the number of people, the staff reported that they learned a lot from CATCH and that their experience with the program was a positive one.

During the interviews, one CATCH staff member mentioned that there were shortages at times in the nursing staff from Livingston, a home health company with which the awardee contracted; the nurses helped with enrollment assessments and visits to participants in their homes. As the pool of participants grew, the number of Livingston nurses assigned to the program actually decreased, which led to some delays in performing enrollment assessments and, later in the program, home visits. Of the nine staff members who responded to a survey question about staffing, most believed that the number of staff relative to the amount of work was a barrier to effective service delivery, though only one person reported that this was a major barrier.

CATCH staff valued training. Most of the staff surveyed reported that they learned new skills for their role in CATCH. The same number of respondents reported that the training helped them to improve their job performance in the program. Three of 10 respondents wished they had been offered more training.

Engagement of providers. The awardee was successful in engaging providers both within the county system and with outside partners. Because provider training was one of the main goals of the intervention, CATCH staff continually engaged providers throughout the cooperative agreement. Engagement was successful due, in part, to the popularity of the spirometers. In addition, the CATCH leadership team reported that once trained, providers would adopt the GOLD guidelines regardless of CATCHpay because they understood them to be the best standard of care for the treatment of COPD. It is possible that the integration of CATCH staging with Ventura County Health Care's Cerner EHR also led to more patients receiving the GOLD standard of care because it was prescribed by default through the EHR system. That said, clinicians retained the ability to customize the care plan as desired.

For the most part, clinicians found that engaging with CATCH was worthwhile, considering the effort required on their part and the gains for their patients. When surveyed, they reported that CATCH did not consume a large portion of their time. Eighty-two percent reported that they spent less than 10 percent of their week on any aspect of CATCH; 61 percent reported that their responsibilities did not change when CATCH began. Sixty percent of clinicians thought that CATCH was worth the effort, with only 7 percent reporting that it was not worth the effort.

By Year 3, more providers were recruited as the CATCH program had gained visibility in the county through word of mouth. In addition, the Year 2 changes to EHR software at the Ventura County Health Care Agency ensured that providers were prompted to enroll participants in CATCH as they entered the diagnosis codes into the system.

Engagement of program participants. The awardee was successful in engaging program participants. CATCH staff maintained contact at least once a month with all participants via text message. The messages included questions such as "Do you remember that you have COPD?" or "Are you taking your medications?" This prompted recipients to reply, "Yes" or "No." The messages also included the CATCH phone number, allowing participants to call staff if they had any questions. Participants with more advanced stages of COPD were seen at the CATCH clinic every few months. Eleven of the 12 CATCH staff who responded to the survey believed that the program had successfully engaged participants.

c. Barriers and facilitators associated with service delivery effectiveness

Ventura County Health Care staff mentioned one barrier to service delivery effectiveness: being short staffed, especially among the nurses who conducted the initial home visits. The awardee's vision of providing each participant with a home visit was ambitious. Doing so required a nurse to travel across a large county and to each participant's home for a visit that could take more than an hour. As mentioned, the awardee contracted with Livingston to perform some of the assessments in the initial home visit, but there were not always enough nurses to do so in a timely fashion.

CATCH staff mentioned two facilitators of service delivery effectiveness: (1) the CATCH clinic and (2) the program's adaptability and the process for customizing the plan of care for each participant.

First, the completion of the CATCH clinic in Year 2 created a space for participants to receive specialized services. Although the program emphasizes home visits, some services such as the 6 minute walk, had to be administered in a clinical setting. Additionally, CATCH staff learned that late stage COPD patients were waiting 3-6 months to see the pulmonologist. The clinic allowed staff to hasten the treatment of advanced COPD patients and to provide pulmonary rehabilitation, nutrition education and counseling for participants with more severe stages of COPD.

Second, one CATCH staff member praised the program's adaptability, especially the process for customizing a care plan for each participant. Each week, CATCH staff met to discuss each new participant and his or her plan of care. The plan focused not only on medication and services directly related to COPD but also—and more comprehensively—on each individual's needs, including behavioral health and everyday needs such as food and housing.

C. Assessment of perceived program effects on the delivery of care and outcomes

Participating clinicians and CATCH program staff reported that the program had a positive effect on the delivery of care. Provider education led to an understanding of the evidence based GOLD Guidelines, spirometry testing, COPD staging, and appropriate levels of care. CATCH staff reported improved outcomes from a decrease in ED visits to an increase in the proportion of spirometry tests administered system wide. In addition, the home visits allowed participants to receive services and supports where they lived.

Among clinicians surveyed, 84 percent reported feeling that CATCH was somewhat or very effective in achieving its goals. The most commonly cited reasons for this perception were improved coordination between care providers and sites (42 percent) and better treatment plans and outcomes for participants (23 percent). All of the CATCH staff surveyed agreed that the program was making a difference in meeting critical needs in the community and that CATCH had been effective in meeting its goals.

During our interviews with CATCH staff, they said that the program made providers more aware of COPD and built a better understanding of the GOLD Guidelines and PFTs. In addition, the staff found that participants were more likely to recognize symptoms early, to take their medication as directed, and to use respiratory equipment correctly.

Based on analyses conducted by the awardee, health outcomes improved over the span of the cooperative agreement. The awardee reported to the implementation and monitoring contractor that exacerbations, hospital visits, and ED visits decreased; and the total cost of care of COPD patients went down 5 percent, from $1,844 per beneficiary per month before CATCH to $1,750 after Year 3. These data were gathered by the awardee using the Ventura County Health Care EHR system.

D. Implications of implementation findings for the design and interpretation of an impact analysis

The Ventura County Health Care leaders as well as CATCH staff perceived that the program had a positive impact on participants. However, some implementation issues may affect the interpretation and analysis of this impact.

First, Ventura County Health Care included patients with COPD and patients at risk for COPD. This outreach component of the program will make it harder for the impact analysis team to develop a control group. There is no way to identify people at risk for COPD by using claims data in neighboring counties because this population will not have identifiable encounters with health care providers.

Second, the awardee is closing the program down in different ways for different participants. CATCH staff tried to link participants with high needs with other care coordination services in the county. Some participants may benefit beyond the end of the cooperative agreement in different ways depending on the care services they were connected with during the program.

Finally, the impact team may not receive data for the most recent CATCH participants as the program closes down. As a result, there may not be enough data to measure the impact of interventions made later in the program. For example, as explained in Chapter V, the CATCH clinic will continue to exist as a pulmonary rehabilitation clinic. However, the staff was reduced, and the Registered Dietitian (RD), Behavior Health Clinician and the RRT will no longer provide services there. It may be difficult to trace the effects of the program itself during the cooperative agreement with the residual effects after the end of the program.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Ventura County Health Care's CATCH program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Ventura County Health Care's CATCH program ended August 30, 2017. The projected sample sizes in Table III.1 are based upon actual enrollment as of February 28, 2017 and are the maximum number of beneficiaries that could be included in our evaluation to allow for all participants to receive six months of program exposure, a requirement for inclusion in our evaluation to minimize dilution of program effects due to limited exposure to the program components. Due to processing lags in Medicaid data, we have not confirmed that all 1,318 Medicaid beneficiaries meet program eligibility for inclusion in our impact evaluation.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: Ventura County Health Care

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure as of February 28, 2017	517[a]
Projected Medicaid population with 6 months of program exposure as of February 28, 2017	1,318[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	2,219
Likelihood of all-cause hospitalizations	1,885
MDE sample size requirement to detect 20% effect	
Total expenditures	555
Likelihood of all-cause hospitalizations	471
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Questionable, moderate rate of identification of treatment group with claims with some dilution of treatment effect within an intent-to-treat framework

Table III.1 *(continued)*

Evaluability domain	Response
Likelihood of solid comparison group	Some issues, but probably surmountable; expect to select a comparison group
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff, clinician, and beneficiary surveys
Implementation data that will be analyzed	None

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We expect to conduct a rigorous impact evaluation by using difference-in-differences estimation with matched comparison groups of Medicare and Medicaid adults who have a diagnosis of COPD if we can pool the two populations. We will derive comparison groups that are similar in terms of medical, payer, and demographic characteristics. We will select a propensity score matched comparison group of beneficiaries residing in two counties north of Ventura County—San Luis Obispo and Santa Barbara counties—by using their claims history to match the patterns of COPD-related diagnoses and care observed in claims for the treatment group. A significant challenge for the evaluation is matching the comparison group with the treatment group on the stage of COPD. Currently, about 20 percent of the treatment group does not have a COPD stage in the data provided by the awardee. For most without a stage, the data indicate that the test for staging is pending. Further, nearly one-half of the treatment beneficiaries do not have COPD but are at risk of developing it. Although ICD-10 coding will help with matching based upon COPD stage, it may not be possible to develop a solid comparison group of beneficiaries matched on the risk of COPD. The number of participants who enrolled before March 2017, and who had a stage indicating COPD produces a sample that is sufficient to detect a 20 percent effect on total expenditures and on all-cause hospitalizations. There may be challenges, however, with matching those who are newly diagnosed with COPD, as their medical histories may be incomplete in claims data. If those without a prior COPD diagnosis cannot be matched, then the sample size may fall below the MDE sample sizes required to detect a 20 percent effect.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents baseline characteristics of the treatment group, which we measured for the year prior to enrollment in the CATCH program. The treatment group consists of participants who have been diagnosed with COPD or who are at risk of contracting COPD. Patients agreeing to participate in the program are given a pulmonary function test using a spirometer, with the results of the test indicating whether the participant has COPD and the severity of the disease if diagnosed. Those who are diagnosed with COPD are classified into one of five stages of severity that are used by providers to identify strategies for care, which are aligned with each stage. Participants deemed at risk of COPD are enrolled in smoking cessation programs and receive case management services to mitigate their COPD risks.

Below we summarize beneficiary characteristics, core claims-based outcomes, and awardee-specific claims-based outcomes during a one-year baseline period for Medicare FFS

beneficiaries participating in the CATCH program. Ventura County Health Care began enrolling participants in the CATCH program in January 2015. As of May 31, 2016, 616 Medicare FFS beneficiaries had enrolled. In addition, 779 beneficiaries, who were predominantly Medicaid managed care beneficiaries, were enrolled in the CATCH program over this period.[3] Subsequent analysis will include Medicaid managed care beneficiaries once the encounter data become available.

For the purpose of presenting baseline characteristics in this report, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before enrollment). In addition, treatment group beneficiaries had to enroll in the awardee's program on or before May 31, 2016, in order to ensure a sufficient run-out period to capture nearly all claims for the most recent participants. The calendar period covered by the baseline quarters is based on the enrollment date for each participant and therefore varies by participant. The enrollment date is defined as the date on which Ventura County Health Care received a beneficiary's consent to enter the program. After we excluded individuals who did not meet the above criteria, a total of 416 participants were included in the analysis of baseline characteristics for this report.

The CATCH program comprises a group of high-need participants (Table III.2). Just over half of them (51 percent) are younger than 65, and more than half (56 percent) are male. Three-quarters of the participants are white. Hispanics represent the next-largest ethnic group at 13 percent. Eighty percent of the FFS participants are Medicare and Medicaid dual eligibles—more than four times the share in the general Medicare FFS population (18 percent)—which indicates a high level of economic need. CATCH participants also have substantial health care needs. The FFS participants are twice as likely as their counterparts in the general Medicare population to have become eligible for Medicare because of a disability (66 percent versus 34 percent, respectively). Furthermore, the average hierarchical condition categories (HCC) risk score for CATCH Medicare FFS beneficiaries is 58 percent above the average score for Medicare FFS beneficiaries nationwide.

[3] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in Ventura County Health Care's program through May 31, 2016

	All participants (N = 416)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	214	51
65 to 74	125	30
75 to 84	58	14
85 and older	19	5
Gender		
Female	182	44
Male	234	56
Race		
White	316	76
Black	14	3
American Indian, Alaska Native, Asian/Pacific Island American, or other	25	6
Hispanic	58	14
Original reason for Medicare eligibility		
Old age and survivor's insurance	141	34
Disability insurance benefits	273	66
End-stage renal disease (ESRD)[a]	2	0.48
Hospice[b]	1	0.24
Medicare/Medicaid dual status, percent dual[b]	331	80
HCC score[c]		Statistic
Mean		1.58
25th percentile		0.69
Median		1.18
75th percentile		2.11

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the beneficiary consented to participate in the program. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Table III.3 shows the HCC COPD diagnosis (HCC 111) from claims over the baseline year and COPD stage for those participants for whom a stage has been assigned.[4] The COPD stage indicates the severity of the condition, with COPD stages 0 through 4 assigned from the result of a spirometry test. The presence of a COPD diagnosis in the baseline year can distinguish previously diagnosed participants from those who were diagnosed with COPD as a result of the program. The diagnosis is critical for identifying a comparison group because participants will be matched on the basis of having a COPD diagnosis and comorbidities that we find to be related to the COPD stages. Two-thirds of participants (66 percent) have been assigned a COPD stage 1-4, with the remaining third at risk of COPD (stage 0). Through smoking cessation programs and case management, the CATCH program hopes to prevent COPD for at-risk participants. For those with a spirometry test indicating COPD, one-half have low or moderate severity COPD (stages 1 and 2). Of these participants, just over half (54 percent) had an HCC COPD diagnosis prior to enrollment, with the remainder appearing to be newly diagnosed. Those having severe or very severe COPD (stages 3 and 4), on the other hand, predominantly show a history of COPD, with 86 percent having a COPD diagnosis prior to enrollment.

The substantial health care needs of CATCH participants are reflected in the high level of their health care expenditures and utilization. Table III.4 shows a common set of utilization and cost measures, including core measures from the Center for Medicare & Medicaid Innovation (CMMI). Because COPD exacerbations often lead to emergency department (ED) visits and inpatient stays, Ventura County Health Care is seeking to lower the risk of exacerbations through improved care under the CATCH program and, in turn, to lower health care utilization among COPD patients. We examined the baseline cost of care by calculating average per beneficiary per month (PBPM)[5] Medicare expenditures in total and by major types of services. The total average PBPM Medicare expenditure during the baseline year was $1,262—about 60 percent above the 2014 national average for Medicare FFS beneficiaries of $792.[6] The average PBPM Medicare expenditure in the baseline year ranged from an average of $851 in the first quarter to $1,498 in the fourth quarter. Over the four quarters, there was a general increase in the intensity of services. Medicare expenditures for inpatient services ($542 PBPM) were the largest driver of the total cost of care for participants, followed by expenditures for physician services ($288 PBPM) and outpatient services ($186 PBPM).

[4] All participants were given a spirometry test and a COPD stage, but both can be delayed.

[5] Months referred to in our calculations are 30-day periods rather than calendar months.

[6] Except for ambulatory observation stays, the national cost and utilization data here and in the next paragraph are from the Centers for Medicare & Medicaid Services, "Public Use File. New Data on Geographic Variation." Available at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

Table III.3. COPD stage by baseline year COPD diagnosis (HCC 111), Ventura County Health Care

	HCC 111 (chronic obstructive pulmonary disease) = 0				HCC 111 (chronic obstructive pulmonary disease) = 1				Total frequency	Total percentage
	Frequency	Percentage	Percentage of stage	Percentage of HCC category	Frequency	Percentage	Percentage of stage	Percentage of HCC category		
Stage										
0	46	24.21	71.88	54.76	18	9.47	28.13	16.98	64	33.68
1	8	4.21	53.33	9.52	7	3.68	46.67	6.60	15	7.89
2	21	11.05	43.75	25.00	27	14.21	56.25	25.47	48	25.26
3	6	3.16	13.04	7.14	40	21.05	86.96	37.74	46	24.21
4	3	1.58	17.65	3.57	14	7.37	82.35	13.21	17	8.95

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Note: The HCC software used to create the baseline HCC score is used to identify participants with COPD diagnoses in the baseline year. This indicator uses HCC logic for classifying COPD but does not use the hierarchical component, which would limit the classification to those without any higher order conditions.

COPD = chronic obstructive pulmonary disease; HCC = hierarchical condition category.

Table III.4. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in Ventura County Health Care's CATCH program through May 31, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	416	403	407	415	416
Average Medicare expenditures PBPM[a]					
Total	1,262 (108)	851 (97)	1,487 (275)	1,199 (147)	1,498 (148)
Acute inpatient	542 (69)	286 (60)	708 (192)	530 (99)	635 (93)
Inpatient other[b]	57 (19)	60 (30)	130 (71)	8 (6)	32 (12)
Outpatient[c]	186 (23)	153 (19)	204 (37)	163 (24)	225 (30)
Physician services	288 (24)	242 (23)	302 (41)	293 (37)	315 (28)
Home health	78 (10)	52 (12)	42 (11)	81 (16)	136 (22)
Skilled nursing facility	86 (20)	38 (21)	81 (34)	97 (40)	127 (45)
Hospice	1 (1)	1 (1)	0 (0)	0 (0)	5 (5)
Durable medical equipment	22 (3)	19 (4)	19 (4)	26 (5)	25 (4)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	518 (57)	321 (67)	514 (82)	537 (94)	692 (93)
Outpatient ED visits	1,112 (101)	873 (125)	1,067 (150)	1,093 (154)	1,404 (178)
Observation stays	128 (23)	100 (40)	109 (35)	98 (30)	202 (47)
Primary care visits in any setting	7,466 (335)	5,952 (332)	6,767 (416)	7,728 (589)	9,337 (528)
Primary care visits in ambulatory settings	5,777 (229)	5,260 (290)	5,216 (299)	5,757 (317)	6,837 (337)
Specialist visits in any setting	9,420 (572)	8,903 (635)	8,724 (694)	9,279 (797)	10,731 (768)
Specialist visits in ambulatory settings	7,233 (435)	7,247 (518)	6,501 (459)	6,733 (510)	8,423 (577)

Table III.4 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Measures of any health care utilization					
Percentage with a hospital admission[d]	28 (2)	7 (1)	10 (2)	9 (1)	14 (2)
Percentage with an outpatient ED visit[e]	43 (2)	15 (2)	17 (2)	16 (2)	21 (2)
Percentage with an observation stay[f]	10 (1)	2 (1)	2 (1)	2 (1)	5 (1)
Percentage with a 30-day readmission among all discharges	18 (3)	11 (6)	17 (6)	22 (6)	18 (5)
Percentage of participants with a readmission among all participants	5 (1)	1 (< 0.5)	1 (1)	2 (1)	2 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures were calculated from all claims for each participant with at least one eligible day during the baseline year or any baseline quarter. These measures are means for the individuals in the analysis sample, not institutional means.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter. This measure is not a Center for Medicare & Medicaid Innovation priority measure for monitoring and evaluation.

[e]The percentages shown do not include ED visits that resulted in an inpatient admission.

[f]The percentages shown do not include observation stays that resulted in an inpatient admission.

ED = emergency department; PBPM = per beneficiary per month.

In the baseline year, the annual inpatient hospitalization rate of 518 per 1,000 beneficiaries was well above the national rate of 274 inpatient hospitalizations per 1,000 beneficiaries in 2014. Eighteen percent of participants were readmitted within 30 days of being discharged, nearly identical to the national rate for all Medicare FFS inpatients. The CATCH program specifically targets this outcome by requiring participants to be contacted immediately by a staff registered nurse or respiratory therapist after a discharge and to visit a primary care provider within seven days. In the baseline year, the rate of 1,112 ED visits that did not lead to a hospitalization per 1,000 participants was two and a half times the 2014 national Medicare FFS rate of 454 per

1,000 beneficiaries. The annual observation stay rate of 128 per 1,000 participants was just over double the observation stay rate of all Medicare FFS beneficiaries in 2013 (60 per 1,000 beneficiaries).[7] Twenty-eight percent of CATCH participants had at least one hospitalization during the year before enrollment, 43 percent had an ambulatory ED visit at least once, and 10 percent had at least one observation stay. Other than inpatient and outpatient utilization, CATCH participants were more likely to visit specialists than primary care providers—9,420 specialty visits per year per 1,000 participants compared with 7,466 primary care visits per year per 1,000 beneficiaries. Through improved care coordination, Ventura County Health Care expects to see a decline in the use of specialty services and a potential increase in the use of primary care. Over the course of the four baseline quarters leading to enrollment, we observed an upward trend in average PBPM total payment and average PBPM payments for inpatient, outpatient, and physician services. We observed a similar trend in rates of hospitalizations, outpatient ED visits, observation stays, and specialty services. These trends indicate that beneficiaries targeted for CATCH were high-cost, frequent users of acute care during the year before enrollment but that they were also extremely high-cost, frequent users of acute care services in the quarter immediately before enrollment. This trend has important implications for selecting a comparable cohort of Medicare FFS beneficiaries as the comparison group.

COPD symptoms can lead to inpatient stays and ED visits. Table III.5 presents rates of both for patients with a principal diagnosis of COPD or a principal diagnosis of respiratory failure and a secondary diagnosis of COPD. With an annualized rate of 103 stays per 1,000 beneficiaries, COPD-caused inpatient stays constitute about one-fifth of all inpatient stays for CATCH participants. COPD-caused ED visits made up a smaller share of all ED visits for participants but still represented a substantial level of utilization, given the rate of 160 visits per 1,000 enrollees. The CATCH program is expected to reduce the likelihood of exacerbations and, in turn, of ED and inpatient utilization.

Table III.5. Measures specific to Ventura County Health Care

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Q1 (10 to 12 months before enrollment)	Q2 (7 to 9 months before enrollment)	Q3 (4 to 6 months before enrollment)	Q4 (1 to 3 months before enrollment)
Total number of enrollees	416	403	407	415	416
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions with principal diagnosis of COPD or respiratory failure with COPD	103 (22.41)	40 (19.83)	79 (30.93)	98 (40.85)	192 (50.06)
Outpatient ED visits with principal diagnosis of COPD or respiratory failure with COPD	160 (33.08)	151 (53.14)	168 (46.54)	127 (34.4)	192 (56.99)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016.

COPD = chronic obstructive pulmonary disease; ED = emergency department.

[7] See MedPAC, "A Data Book: Health Care Spending and the Medicare Program," June 2015.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

The awardee's payment model combined a discounted bundled payment for treating COPD patients with an incentive payment for PCPs to follow evidence-based COPD clinical guidelines. The evidence-based clinical guidelines are still in use today, as is the bundled care plan in the EHR. However, CMS did not adopt the bundled payment and the provider incentive funding ended with the HCIA grant funding.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Ventura County Health Care initially proposed a pay-for-performance payment model (CATCHpay) for the target population: Medicare and MediCal patients within Ventura County's health system who have COPD. The model combined a discounted bundled payment for treating COPD patients with an incentive payment for PCPs to follow evidence-based COPD clinical guidelines. The bundle of services proposed under the payment included 12 months of care using the cost categories of inpatient/outpatient hospital costs, emergency services, professional primary care and specialist care, diagnostic imaging and X-ray, laboratory services, home health care, covered nonprescription services, and outpatient (Part D) and professionally administered (Part B) drug therapies. The bundled payment component of the payment model was not approved by CMS. As a result, the awardee implemented the incentive payment component of the model and embedded the bundled treatment plan into the EHR but continues to bill fee for service rates.

Participating PCPs had to meet certain requirements to receive incentive payments: (1) complete a CATCH CME certification course and follow best practices (including the GOLD guidelines) and (2) use the PowerForm tool in the EHR. The COPD PowerForm was located in the Cerner EHR system that is accessible to all providers (and linked to the bundled service). The PowerForm acted as an indicator for Ventura County Health Care to easily identify which PCPs would receive an incentive payment and it is also the source of all medical billing. CATCH staff looked at the list of PCPs who followed the guidelines for each patient and cross-referenced their names with the names on the list of PCPs who completed CATCH CME certification. If a PCP met both criteria, he or she received an incentive payment.

Ventura County Health Care paid PCPs every six months. The awardee assigned a set amount of funds for each round of incentive payments. The first month used to identify services rendered was March 2016. Each PCP received a share of the funds depending on the frequency with which they followed the GOLD guidelines for eligible patients in the EHR tool. The more frequently a PCP followed the guidelines, the higher his or her share of the incentive payment.

C. Status of the payment model

Although the provider incentives were successful, the awardee never gained traction on the bundled payment component of the payment model. And although the awardee initially believed that it would receive approval from CMS, it decided not to move forward with the bundled payment based on the lack of interest and response from CMS about this component. Ventura County Health Care therefore focused on using CATCHpay to incentivize PCPs to follow the GOLD guidelines for COPD care.

D. Factors associated with the development of the payment model

During our discussions with Ventura County Health Care, the awardee identified the primary factor that facilitated the design of the payment model: the right stakeholders were involved in discussions as the model was being developed. The stakeholders included a Cerner analyst, the CATCH medical director, and the specialty medical director, who developed the PowerPlan and the PowerForm. The PowerForm became the basis of the bundled payment model and the CATCHpay incentive report.

The awardee faced two challenges while trying to implement the payment model. The first and biggest challenge reported by the awardee was the lack of interest in the bundled payment component. Ventura County Health Care attended several HCIA R2 payment model workgroup meetings in which the discounted bundled payment was discussed, but the awardee reported that CMS did not express an interest in this approach. The awardee mentioned receiving feedback from the CMS consultants who questioned why it wanted to offer the same or better treatment for less reimbursement. The second challenge the awardee faced was, as it moved forward with the incentive payments for providers, it implemented a new financial system and ran into a few delays with sending payments to providers. Specifically, the awardee's Accounting and Contracting department had questions about the payments, which eventually held up the distribution of the payments. The awardee addressed the Accounting and Contracting department's concerns and eventually paid the providers. The awardee also noted that providers were understanding about the delay, and it did not affect their participation in the program.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Ventura County Health Care made progress in sustaining parts of the program funded by the cooperative agreement. The awardee finished integrating GOLD guidelines into the Cerner EHR, which will remain in place beyond the cooperative agreement. Existing clinics will absorb program participants, although they may not receive the same services. Finally, the awardee reported that it would sustain the CATCH clinic and, as of this writing, was awaiting certification to provide pulmonary rehabilitation services there. The awardee decided not to replicate or scale its program in the third program year.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Ventura County Health Care had finalized agreements with Ventura County's Medicaid health insurance plan (Gold Coast) to cover pulmonary services. The awardee had also finalized plans for the CATCH clinic, whose services are a Medicare-covered benefit. The awardee said that it intended to use internal funding to sustain some program positions after the cooperative agreement ended.

C. Implementing the SSR plan: progress and changes

Sustainability. As of the third program year, Ventura County Health Care leaders planned to sustain only certain aspects of the program. First, the awardee finished integrating the GOLD COPD stages in its Cerner EHR system. Going forward, providers selecting COPD diagnoses will be prompted to enter COPD stage information obtained from a spirometry test, and based on the stage, follow-up care recommendations will be automatically populated in the system. Providers will have the option to override the recommendations. Second, the awardee reported that program participants would be absorbed into existing clinics, although the services they receive at those clinics would not be exactly the same as the program services. For example, the awardee leaders doubted that nurses and RRTs would continue making home visits or phone calls to patients. One leader reported that after the cooperative agreement, the CATCH clinic would morph into pulmonary rehabilitation clinic.

In the 12th program quarter, the awardee reported that the CATCH clinic was waiting to be certified for pulmonary rehabilitation by the American Association of Cardiovascular and Pulmonary Rehabilitation (AACVPR). The awardee reported that this pulmonary rehabilitation program, which will provide outpatient therapy services, will be the legacy of the CATCH clinic. As part of this program, the awardee said that it created new PowerCharts (questionnaires used in pulmonary rehabilitation), based on recommendations from AACVPR.

Scalability. Ventura County Health Care did not report that it took steps to scale its program.

Replicability. The awardee did not report that it took steps to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

Integrating the GOLD guidelines into the Cerner EHR system facilitated sustainability, but it was impeded by California's long delay in reviewing the licenses for the pulmonary rehabilitation clinic. The integration of the GOLD Guidelines into the Cerner system will ensure that Ventura County Health Care providers continue to be aware of the GOLD standard of care after the end of the cooperative agreement.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

The end of the cooperative agreement signals the end of primary data collection. Once the awardee submits its final report to the implementation and monitoring contractor, we will review it to determine whether there is new information that provides additional insight into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Ventura County Health Care's CATCH program. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicare beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Village Center for Care

April 18, 2018

Ellen Wilson (RTI International)
Noah Triplett (RTI International)
Cordon Newhart (RTI International)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Village Center for Care (VillageCare) received a four-month, no-cost extension. Enrollment for the Rango program ended on February 28, 2017, and services ended on December 31, 2017. VillageCare used the no-cost extension to provide services to participants through the end of the calendar year, to continue analyses of evaluation data, and to pursue opportunities for ongoing funding.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

Village Center for Care (VillageCare), a community-based nonprofit organization in New York City, used funding from HCIA R2 to create and implement the Rango program. Rango provides support to HIV-positive participants via an integrated mobile platform (Rango.net) and a mobile app. Rango is designed to help people who are HIV-positive take their antiretroviral medications every day exactly as prescribed to maintain their health, which can be challenging for many people. Rango uses an innovative technological approach to support treatment adherence. Key program characteristics are shown in Table I.1. In its HCIA R2 application, VillageCare proposed that Rango be a tool to support care management. However, at the beginning of the cooperative agreement, the awardee changed its focus to instead promote participant engagement in care and disease self-management (for example, taking medication consistently, exercising, managing stress, and healthy eating). Rango promoted disease self-management through a variety of features, including a community forum where participants could post comments and interact with each other, articles with useful information for participants, and medication and appointment reminders. Participants were enrolled in the program for a period of 12 months. After the 12 months, they could continue using most of Rango's services, but they would no longer receive incentives or appointment and medication reminders via text message (although, they could receive reminders via the Rango app).

VillageCare launched the Rango program on April 1, 2015, with the goal of enrolling and engaging 5,036 participants on its platform by the end of the three-year cooperative agreement. VillageCare adapted Rango throughout the three years in response to user feedback—for example, by dropping some features that were underutilized and adding others. The awardee's expected outcomes for Rango included (1) increasing participants' retention in HIV/AIDS care; (2) supporting treatment adherence; (3) increasing participants' time in first-line (that is, least burdensome and least costly) HIV/AIDS treatment; and (4) reducing the costly hospitalizations and outpatient services that are associated with treatment failure.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	To improve adherence to HIV/AIDS treatment through the use of an integrated mobile platform and app
Major innovation	Use of innovative technological approach to improve adherence to HIV/AIDS treatment
Program components	• **Patient engagement.** Rango supports self-management of HIV/AIDS and associated conditions • **Health information technology.** Mobile platform and app with educational, motivational, and reminder features
Target population	Participants age 18 and older who are diagnosed with and prescribed medication for HIV/AIDS; live in New York City or the surrounding areas; and are covered by Medicaid, Medicare, or both (dual eligible). Most current participants are Medicaid only.
Theory of change/ theory of action	Participants' use of electronic self-care tools will improve their adherence to HIV treatment and their engagement in and satisfaction with their care. VillageCare anticipates that Rango will reduce the costs associated with treatment failure and eliminate the need for more burdensome and expensive therapies.

Table I.1 *(continued)*

Program characteristic	Description
Payment model	Capitated payment
Award amount	$7,983,297
Effective launch date	4/1/2015
Program setting	• Recruitment at partnering community-based organizations, managed care organization sites, and primary care provider locations • Services delivered through Rango platform
Market area	Urban
Market location	New York City and surrounding areas—including the Bronx, Brooklyn, Manhattan, Queens, and Staten Island, as well as Nassau, Suffolk, and Westchester counties
Target outcomes	• Adherence to treatment • Participant engagement • Participant satisfaction

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on six factors. First, the awardee enrolled 4,367 participants—87 percent of its enrollment target—by the end of the cooperative agreement. Second, the awardee developed and launched the Rango platform and mobile-enabled website on schedule. Third, the awardee hired the staff necessary to implement the program and provided them with effective training, despite problems with turnover in the program liaison position. Fourth, the awardee engaged multiple referral partners to help with recruitment and continued to add additional partners as needed. However, the level of support from staff at the partner organizations varied. Fifth, the awardee had relatively low rates of disenrollment among participants, although the level of engagement among participants varied. Finally, participating clinicians and other implementation staff reported that the program had a positive effect on some participants' health behaviors. However, they also reported that some of the most challenging participants were unaffected.

Impact evaluation. Although we do plan to carry out a rigorous assessment of the impacts of VillageCare's Rango program, the analysis is still in progress and not included in this report.

Payment model. VillageCare was still developing its payment model at the end of the third year of implementation. The awardee planned to continue discussions with the managed care organizations (MCOs) that provided patient referrals for this HCIA R2 demonstration. However, VillageCare was not in active negotiations for payment.

Sustainability plans. In an effort to sustain the Rango program, VillageCare worked to gain approval from its two MCO partners for the payment model. If Rango showed evidence of effectiveness and the price was low enough, one MCO was interested in providing payment as

well as expanding the program to its other members. The awardee also was trying to obtain grants from federal agencies and philanthropic foundations, and to potentially scale the program to other providers and populations. If VillageCare could not secure funding during its four-month, no-cost extension, the awardee said that it would have to discontinue the program.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to a survey of non-clinician staff on their perceptions of the program's effect on care delivery. The non-clinician staff survey, which was fielded from July 2016 to October 2016, achieved a response rate of 91 percent. The sample size was 11 respondents. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we concluded that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

VillageCare partnered with MCOs, primary care providers, and community-based organizations that work with HIV-positive populations in New York City to help with participant recruitment. In Year 1, VillageCare established contracts with eight referring partners. In subsequent years, the awardee added additional, noncontractual referring partners. Participants were referred to Rango by staff at partner organizations or through word of mouth. In addition, program liaisons (Rango staff responsible for recruitment and enrollment) spent time at the partner organizations to recruit participants directly. VillageCare trained staff at partner organizations on who was eligible for Rango, what the benefits to clients were, and how to complete referrals.

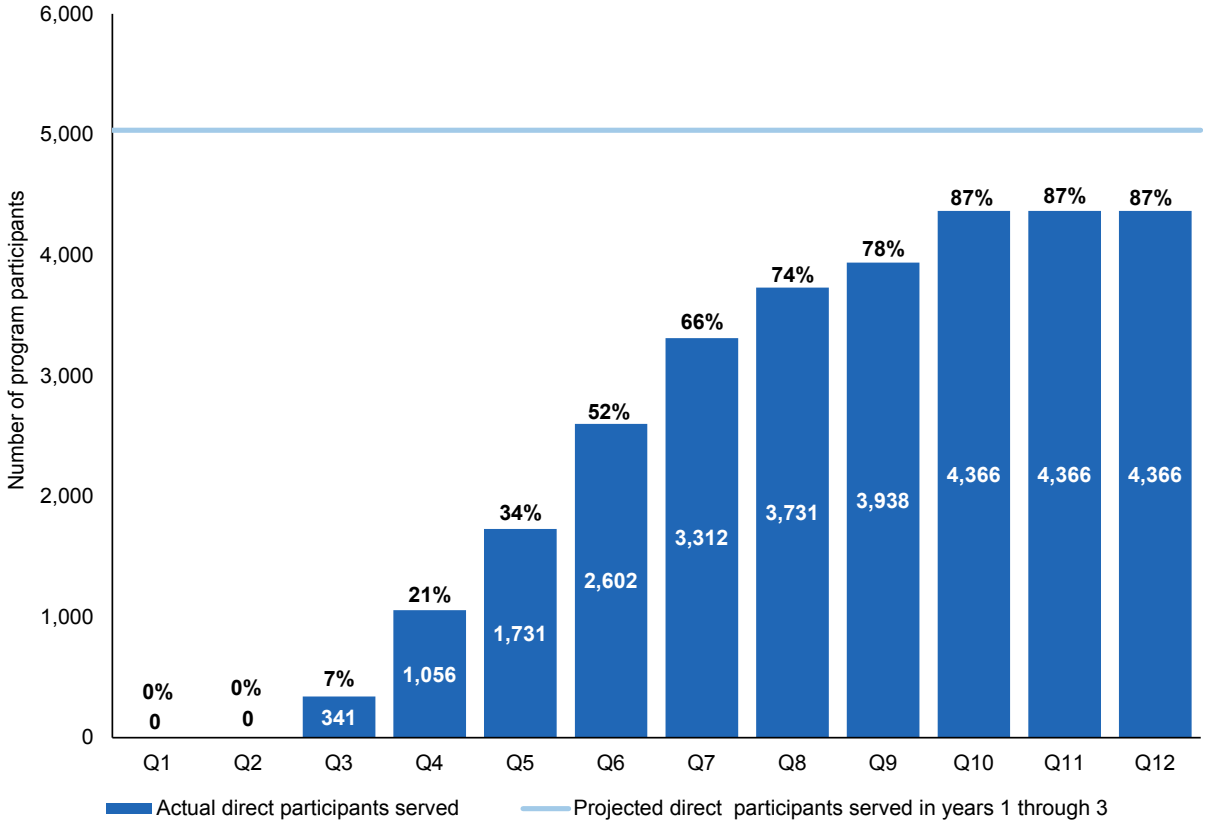
Once potential participants were referred, a program liaison helped them enroll in the program. The enrollment process evolved over time to become more streamlined. Originally, most enrollments were conducted by appointment. During enrollment, the program liaison walked participants through all of the paperwork (informed consent, data sharing authorization, and registration forms) and provided a demonstration of the Rango platform. This process was too time-consuming, however, so VillageCare implemented a variety of changes to make enrollment more efficient. By Year 3, program liaisons conducted most enrollments on-site at referring organizations (for example, in waiting rooms) instead of by appointment. The liaisons also gave participants an enrollment packet that they completed mostly on their own. Program liaisons were available to help participants if they had any questions. They would sometimes do a short demonstration of the platform or help participants download the Rango app, if needed. In addition, participants could still make an appointment for a full demonstration if they wanted one.

VillageCare offered participants a monthly incentive. Originally, the incentive was $40 per month toward their cell phone payment. Early in Year 2, the awardee reduced the amount to $35 per month. In the middle of Year 2, VillageCare changed the incentive to a $35 per month electronic gift card (instead of a $35 contribution toward a cell phone payment).

b. Evidence of enrollment effectiveness

VillageCare originally planned to enroll 5,121 participants but slightly reduced the number to 5,036 at the end of Year 1. VillageCare exceeded its enrollment goals in the first year of the cooperative agreement. That year, VillageCare enrolled 1,056 participants—approximately 8 percent more than the 981 participants it had projected for Year 1. At the end of Year 2, VillageCare was on track with its enrollment. By that time, the awardee had enrolled 3,731 participants—just over its target of 3,728. In Year 3, however, enrollment tapered off and the awardee ended up with somewhat fewer enrollments than it had projected. Overall, VillageCare enrolled 4,366 direct participants from April 2015 (when it launched its program) through February 2017—about 87 percent of its three-year enrollment goal (Figure II.1). Because VillageCare received a no-cost extension through December 31, 2017, nearly all participants had the opportunity to receive the full 12 months of Rango. However, a few hundred participants who were enrolled during the last 2 months of enrollment had the opportunity to receive only 10 or 11 months of services.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. VillageCare changed its projection from 5,121 participants to 5,036 participants at the end of Year 1.

c. Barriers and facilitators associated with enrollment effectiveness

VillageCare's progress in meeting its three-year enrollment goal was influenced by several factors. In Year 3, the awardee had overcome some of the challenges to enrollment it confronted in prior years, but also faced new challenges. Ultimately, it was not quite able to reach its enrollment goal.

As described in previous annual reports, the primary barriers to enrollment that the awardee experienced before Year 3 were (1) inefficiencies in the enrollment process and (2) the fact that some referral sites had already referred most of their eligible patients to Rango. The awardee addressed these challenges by streamlining the enrollment process, adding new recruitment sites at some of its existing referral partners, and adding new partners. A barrier to enrollment that the awardee experienced in all three years was that some eligible participants did not want to enroll. In the staff survey, the reasons most frequently mentioned for why some participants chose not to

enroll were that the technology components seemed too complicated; concerns around privacy/confidentiality; and that the participant did not understand the program.

In Year 3, the awardee encountered some new barriers to enrollment, including (1) a six-month cessation of referrals from one referral partner; (2) a lack of engagement by staff at some referral partners; and (3) several issues related to the fact that the program was nearing an end, including resignations of enrollment staff, reluctance of some referral partners to refer patients to the program given that it was going to be ending soon, and confusion among some enrollment partners about when the enrollment period ended. These barriers are described in more detail below.

Between July 2016 and December 2016, enrollment was suspended at all sites connected with Mount Sinai Health System—one of the awardee's largest referral partners. A Mount Sinai researcher had identified some deviations from proper enrollment procedures for a small number of participants referred from Mount Sinai clinics. As a result, VillageCare had to suspend enrollments at all Mount Sinai sites until a corrective action plan (CAP) was established. After the CAP was in place, the awardee had to schedule meetings with staff at all six Mount Sinai sites to ensure that they understood the new procedures, which caused further delays. Once enrollment activities resumed, awardee leaders said that they had lost a lot of momentum, and it was difficult to regain staff buy-in.

Another barrier was that not all staff at partner organizations were actively supporting the program and helping with referrals. Awardee staff said that in Year 3 they came to a new realization of the importance of having this kind of support. Program liaisons could approach patients and try to enroll them without staff involvement, but they found that they were much more successful if the staff were involved. Staff knew which patients were eligible and which were not, so they could help to screen for ineligibles. In addition, because the patients knew and trusted the referral staff, their endorsement of the program made patients more likely to want to participate. Given this realization, awardee staff increased their efforts to promote the program to staff at the recruitment sites. For example, they used posters and brochures and tried to have more face time with the referral staff. This helped remind the referral staff about Rango and provided an opportunity to clear up any misunderstandings (such as the confusion about when the enrollment period ended). However, program leaders noted that some providers were not active in promoting Rango because they were not convinced that a technology-based intervention would address the complex needs of patients who most needed support in managing their care.

> "I think if the patient does not perceive that they're getting a thumbs-up to participate [from their provider], then that is a real negative."
>
> —Program leader

The fact that the program was nearing an end had several negative effects on enrollment in Year 3. First, some of the program liaisons (the staff responsible for enrolling new participants) were anxious about the fact that their jobs would be ending soon, so they took new jobs elsewhere, leaving the program with a shortage of program liaisons. Second, staff at some of the referral partners were reluctant to make referrals during the final months of the enrollment period because they thought it could be a negative experience for patients to get accustomed to the support provided by Rango and then lose it in a few months when the program ended. Third,

some partners were confused and thought that enrollment had ended before it did, so they stopped referring patients. In hindsight, program leaders said they should have front-loaded enrollment more, to avoid these types of problems.

One interviewee noted, however, that the fact that the program was nearing an end also had some positive effects on enrollment. Some potential participants who had been procrastinating about enrolling were motivated to enroll when they realized it was almost too late. Similarly, some referral partners were also motivated to get their patients connected to the program before it ended.

Program leaders mentioned two additional facilitators to enrollment: (1) the financial incentive to participants and (2) having champions at partner organizations. As noted in the second annual report, the $35 gift card that participants received was a powerful motivator to participants to enroll, even if it was not as compelling as the $35 to $40 cell phone payment that participants had received in Year 1 and the first half of Year 2. Champions at partner organizations helped enrollment by explaining the program to their colleagues and by looking for opportunities to promote Rango. For example, if they knew of a new member event or a new member publication coming out, they could suggest that as an opportunity to showcase Rango. If a new staff person was hired, they would make sure that person was trained on Rango.

2. Delivery of program services

a. Description of and changes to service delivery model

Rango was a completely new program developed under the HCIA R2 cooperative agreement. It was a technology-based program designed to improve adherence to HIV treatment through use of an integrated mobile-enabled website and mobile app. The app was not completed until midway through Year 2, however, so only the website was available until then.

At the time of its HCIA R2 application, VillageCare had envisioned that Rango would focus on care management. After entering into the cooperative agreement, the awardee decided to focus on supporting self-care instead.

When Rango was launched, it included the following features:

- Text reminders to take and refill medications
- Medication tracker
- User profile and avatar
- Online community (discussion boards, friend requests, member search, private messaging)
- Contact forms to request customer service
- Library of self-help articles
- 24-7 virtual support groups on a variety of topics
- Telephone Q and A and messaging with health coaches
- Connecting participants to trained peer mentors

VillageCare planned two phases of development for Rango. After the Phase 1 launch, it reviewed data on program use and participant feedback to determine what changes to make, and then rolled out a revised version of Rango (Phase 2) in the middle of Year 2. The new version dropped some features that had been underutilized (virtual support groups and connections to peer mentors) and added others (live chat with health coaches, a searchable social services database, and enhanced communication features on the home page). After the launch of the revised version of Rango, VillageCare continued to make refinements in collaboration with its technology vendors.

As described in the first annual report, VillageCare changed its staffing plan significantly from what it proposed in its HCIA R2 application, partly in response to the shift in orientation from care management to self-care and in response to needs that emerged during the course of the first year. For example, the awardee decided to hire seven program liaisons instead of a larger number of care managers, and combined multiple part-time health coach positions into two full-time positions. The health coaches were professional staff (one health coach was a licensed social worker and the other was a dietician), while the program liaisons were paraprofessional staff. In total, VillageCare reported 12.26 new hire full-time equivalents funded by the HCIA R2 cooperative agreement, or 82.8 percent of its three-year target.

The two health coaches developed the content for Rango and worked to engage participants by hosting monthly wellness challenges, promoting events and discussions, and encouraging participants to think about their health. Health coaches wrote the articles for the library and interacted with participants through a variety of channels, including discussion boards, live chat sessions, email, and phone calls. Health coaches published new content monthly, which was driven by participants' top questions or concerns.

Rango's theory of action was that referral partners would refer potentially eligible patients to program liaisons for screening and enrollment. Program liaisons would screen potential participants for eligibility and enroll them in Rango, if they were interested. Participants would use the various features of Rango according to their needs and interests. Their use of these features would lead to increased patient engagement and improved adherence to treatment. Most of Rango's features were designed to increase patient engagement. For example, the monthly health challenges were intended to get participants actively involved in improving their health. Rango could improve participants' adherence to treatment directly through medication reminders, as well as more indirectly by providing social support. VillageCare hypothesized that improved adherence to treatment for participants would lead to increased time in first-line HIV/AIDS treatment, fewer health problems, and reduced health care costs.

b. Evidence of service delivery effectiveness

Overall, the awardee was effective in its service delivery. VillageCare developed the Rango platform and mobile-enabled website on schedule, although the development of the mobile app was significantly delayed. Some features of Rango were used much more than others, but overall usage of Rango was relatively high. Interviewees said that a significant subset of participants were not meaningfully engaged in Rango and were only motivated to participate by the incentive. However, they also said that another significant subset of participants highly appreciated Rango. The awardee was successful in hiring and retaining staff except for the

program liaison position: The hiring of program liaisons at the beginning of the program was delayed, and high turnover among program liaisons had a negative impact on enrollment. The awardee provided staff with extensive ongoing training, which enhanced staff morale and effectiveness. VillageCare was able to recruit the referral partners it needed to meet its enrollment targets, including eight contracted referral partners in Year 1 and several noncontractual referral partners in subsequent years.

Delivery of intervention services. VillageCare developed the Rango platform and mobile-enabled website on time, with an ambitious schedule of launching seven months after the start of the cooperative agreement. Development of the mobile app was significantly delayed, however. It was not available until midway through Year 2.

Staffing and training. The awardee was mostly successful in hiring staff and providing them with training. VillageCare did not have any turnover among the health coaches, and the turnover that occurred in management positions led to improvements in program implementation. The one position that the awardee had difficulty filling was the program liaison. In Year 1, identifying and hiring the full component of program liaisons took longer than anticipated. In addition, turnover among program liaisons was a problem throughout the program. As noted in the section on enrollment effectiveness, several program liaisons found new jobs and left the program in Year 3, near the end of the enrollment period. The awardee attempted to fill the program liaison positions for the final months with staff from temporary staffing agencies, but it experienced high levels of turnover among the temporary staff as well.

Staff received a significant amount of training, including two trainings when they were hired: (1) a VillageCare orientation (covering topics such as privacy law and the history of the organization) and (2) an orientation specific to Rango. VillageCare also provided ongoing training in response to needs identified during the course of the program. In Year 2, staff received training about updates to the Rango platform, changes in incentives and enrollment procedures, and customer service and client relations. The awardee also often arranged for staff to participate in trainings related to HIV and other relevant topics being offered around the city. Nearly all staff survey respondents indicated that they received formal training and training through staff meetings, and nearly all agreed with statements that the training they received helped them learn new skills that were helpful for their role within the Rango program and that the training helped them improve their job performance.

Recruitment and engagement of providers. Providers were primarily involved in the program as a source for referrals. In Year 1, VillageCare contracted with eight referral partners, although it experienced unexpected delays due to legal negotiations. It retained all eight partners for the duration of the cooperative agreement and added additional, noncontractual referral partners in subsequent years. Interviewees reported that the level of buy-in by staff at the referral partners varied, with some being very committed and engaged and others less so. Program leaders said they would have liked the partner organizations to provide input on the actual content and programming of Rango, in addition to helping with referrals. This did not occur, despite repeated invitations from the awardee. For example, VillageCare would have liked it if partners had identified topics they wanted the health coaches to focus on or if staff with

particular areas of expertise had conducted a live chat with participants. Rango was also set up so that providers could log in and enter medication and appointment reminders, but they never did.

Engagement of program participants. Retention in Rango was better than VillageCare had expected, given the difficult life circumstances of many HIV-positive people. For example, at the end of the eighth quarter, only 36 percent of participants had disenrolled. The average retention time for those who disenrolled was 129 days. Retention had decreased somewhat by the end of the 12th quarter: 45 percent had disenrolled, with an average retention time of 125 days. Program leaders thought that one factor leading to lower retention near the end of the third year was that participants believed that Rango was ending soon. (VillageCare did not receive notice that its no-cost extension had been approved until shortly before the scheduled end of the program.) In addition, after the 10th quarter, enrollment activities ended at the referral sites, so referral staff may have provided less reinforcement of and support for the program.

On average, usage of Rango was relatively high. As of November 2017, the awardee reported that the 4,367 participants had signed in to Rango a total of 277,700 times (an average of 63 times per participant); posted 32,434 messages (an average of 7.4 times per participant); accessed articles 14,973 times (an average of 3.4 times per participant); and received 1.1 million medication reminders (an average of 251 times per participant).

Participants were more engaged in some features of Rango than others. For example, relatively high proportions used the medication reminders (50 percent via the app and 45 percent via SMS text message), posted messages in the community forum (37 percent via the website and 20 percent via the app), or accessed at least one article (33 percent via the website and 25 percent via the app). Features that participants used less included the social services database (8 percent), appointment reminders (3 percent via the app and 2 percent via the website), and live chat with health coaches (4 percent via the website and 2 percent via the app). The program director noted that participants' level of engagement in the community forum exceeded any benchmarks she could find for community forums in other interventions. VillageCare had initially been concerned about needing to heavily moderate the discussion and enforce community guidelines, but found that participants were very kind and helpful to each other. According to the program director, the forum was a "wildly positive place."

> "Via the forums and via the phone lines, we get rave reviews from our participants—unsolicited, on a regular basis. I have a whole brag book that we've put together [of] our participants telling us how much they love Rango and specifically how it's helped them."
>
> —Program leader

In Years 1 and 2, interviewees speculated about whether participants were motivated to join Rango because they valued the program or because they wanted the monthly incentive. By Year 3, the awardee had a clearer sense of participants' motivations because many had "graduated" by then—they had been enrolled for 12 months, at which point they could continue participating but would no longer receive the incentive. Some participants stopped using Rango when the incentive stopped, but others continued to be actively involved. One interviewee said that he thought that the incentive was important for getting people to sign up in the first place, but once they started using Rango, they discovered that they valued it and they continued using it even after the incentive ended. An interviewee from a partner organization said that he thought

participants were perhaps equally divided—a third who were very engaged in Rango, a third who participated only because of the incentive, and a third who were somewhere in the middle. Some interviewees commented that those participants who struggled the most with self-care were also the least engaged and the most likely to drop out when the incentive ended. Nonetheless, interviewees in all three years said that they had heard from many participants that Rango gave them a support system that helped them cope with their illness. Program staff and leaders said that they had received very positive feedback from participants and that participants gave the program high marks in satisfaction surveys.

c. Barriers and facilitators associated with service delivery effectiveness

Factors that supported service delivery effectiveness were (1) effective communication among team members, and (2) aspects of the program that helped to keep participants engaged, including the Rango platform's engaging nature and ease of use, the support of health coaches, extensive customer service support, and the financial incentive for participation.

Program leaders identified good, systematic communication among team members as being key to the program's success. To develop and implement a brand-new technology intervention in the span of three years required an all-hands effort. Effective communication helped the awardee identify solutions to the stream of questions and challenges that arose.

Several factors helped to keep participants engaged in the program. Program leaders and staff reported that participants found Rango engaging. They particularly liked being able to interact with others anonymously. Interviewees also said that, although some participants struggled with the technology, the program (especially the app) was generally easy to use. Participants generally agreed. In a survey, the awardee asked participants to grade Rango (on a scale of A to F) according to how fun and easy it was to use. Participants gave it a solid A. Program staff and leaders also thought that interaction with the health coaches played a key role in keeping participants engaged. Extensive customer support services helped to overcome some of the challenges participants encountered in using the program. Finally, the $35 per month incentive was a strong motivation for participants to enroll and continue their participation in the program.

Barriers to service delivery effectiveness included (1) difficulties with filling the program liaison position, (2) lack of input from users and providers before designing Rango, and (3) participants' challenging life circumstances and lack of technological know-how.

> "Even starting one full year ago, I would frequently have program liaisons come to me asking: What's gonna happen to the program? What's gonna happen to my job? Will I be laid off? When will I be told? And that really hurt morale, and I think it hurt retention."
>
> —Program leader

Filling the program liaison position was challenging from the beginning because of competition from other organizations for staff with comparable qualifications. Turnover in the position was a problem in Year 2 primarily because of conflicts the program liaisons had with their supervisor (who later left the organization). In Year 3, turnover was due to program liaisons' concerns about the prospect of being unemployed when the enrollment period ended. The project director tried to convince the program liaisons to stay, telling them that when the time

came, she would try to find them another position within the organization, but they felt the possibility of a layoff was too risky for them. The awardee tried to fill the vacant program liaison positions for the final months of the enrollment period through temporary staffing agencies, but personnel with the needed qualifications were difficult to find. The delays in hiring program liaisons had a negative impact on enrollment and the turnover in the position was disruptive to the smooth operation of the program.

Program leaders commented that lack of input from providers and participants about the features they were interested in before the development of Rango was another barrier. The awardee spent time and money developing features that people did not use, and those resources could have been better used if the awardee had received more input in advance. The tight time frame for the development and implementation of the intervention did not allow for such formative research, however. On the other hand, program leaders did think it was helpful that they had planned and budgeted for two stages of development. This allowed them an opportunity to receive feedback from program participants after the initial launch and then modify the program in response.

> "Even though we didn't do a ton of user input and user testing at the beginning, we did at least give ourselves the opportunity to correct midcourse."
>
> —Program leader

The primary difficulty with engaging participants was the challenging nature of the target population. HIV-positive patients in general tend to have major problems in their lives, especially those who are having difficulty managing their care. Related to this, the target population was not very technologically savvy. Some participants struggled to use Rango and required extensive customer support.

Demands for customer service were extremely high during the first two years of the program, which led to frustration among participants (if they could not get a timely response to problems), posed a burden for partner organizations (if participants came back to them for help), and strained human resources at VillageCare. In Year 2, the awardee tried numerous strategies to reduce customer service demand—for example, changing the incentive from a cell phone payment to a gift card and trying to improve communication with participants. By the third program year, program leaders reported that the problem had been resolved.

C. Assessment of perceived program effects on the delivery of care and outcomes

Overall, staff thought that Rango had positive effects on participants. In the survey of program staff, most respondents felt the program had been effective in achieving its goals. In addition, all respondents agreed that Rango (1) was making a difference in meeting a critical need in their community, (2) had a positive impact on the quality of care and services they provided to participants, and (3) had a positive impact on participants' access to care or services and on the achievement of their health goals.

During interviews in Year 3, program leaders noted that although they did not yet have data on long-term outcomes such as a reduction in hospitalizations or ED visits, the data they did have were promising. For example, VillageCare's survey data showed that scores from the patient activation measure (PAM) increased from 62.8 at enrollment to 66.2 at 270 days.

Research on the PAM has shown that a 1-point increase in the PAM led to a 2.8 percent increase in HIV-positive patients' medication adherence. The awardee's survey data also showed that adherence had improved. Program staff commented that many participants had mental health problems in addition to being HIV-positive. By providing a safe, nurturing space, Rango had a positive effect on participants' mental health, which in turn helped empower them to better manage their health. Staff noted, however, that behavior change is a process that takes time and that participants were at different stages of change. The two partners interviewed both thought that the proportion of participants that Rango benefited was relatively small, but they thought that there were definitely some who did benefit. They noted that the circumstances that led HIV-positive people to be nonadherent with their medication were complicated, such as housing issues or spousal abuse. They were not convinced that Rango alone would be able to address these kinds of issues, but they thought it could be a useful tool as part of a broader system of support.

Awardee staff thought that Rango's community forum, which allowed participants to share stories and connect with one another, was a powerful tool. Participants also said that it was very helpful to them. Program leaders noted that some participants posted in the community forum that they had not disclosed their HIV status anywhere else. This could make them more likely to disclose to friends and family, leaders said, which could have an impact on medication adherence because disclosure is strongly linked to adherence.

> "Thank you so much for your support and friendship. I think this is a great study for HIV…. It is a very positive way to keep healthy…. Some of us need someone to talk [to] or express our deepest [feelings]."
>
> —Program participant forum post

When awardee leaders discussed the program's success, they pointed to evaluations commissioned by VillageCare to assess the impact of Rango on health outcomes. As of November 2017, findings from one of the evaluations conducted by researchers at New York University were available. The evaluation examined data for 1,339 Medicaid-enrolled individuals who started using Rango in 2015—approximately 30 percent of the 4,402 participants ever enrolled in Rango. The findings suggested that Rango was effective in improving treatment adherence, reducing hospitalizations, and reducing hospital expenditures.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

VillageCare worked continuously to improve Rango, so participants who enrolled in the latter half of the program may show more positive effects than those who enrolled earlier. The awardee made some particularly significant changes in Year 2, including the release of the app, new features, and the debut of the health coaches' monthly wellness challenges. An interviewee from a partner organization said that he thought that Rango became more appealing to participants later in the program, after it was "cleaned up." In addition, several interviewees said that the app made Rango easier to use. If the improved user experience increased the frequency with which participants used Rango or the new features made Rango more effective, these changes may have led to improved participant outcomes.

Program leaders and partners thought the time frame of the project was too short to see significant changes in outcomes such as hospitalizations. They thought that participants needed more time using the services to receive the full benefit, and that it would take more than a year or two for any behavioral changes to manifest as changes in health outcomes and health care costs. They did, however, think that it might be possible to detect changes in behavioral outcomes such as treatment adherence within the time frame.

> "I think that if there was more time for the program and there was more time for people to really use the services, [then] I think that it would show a better outcome in the future."
>
> —Program leader

> "In terms of medication adherence or appointment adherence, I think it's likely that [Rango] is having some impact for some patients. It may be hard to show, though, because most of our patients have suppressed viral loads without any extra intervention. You're really trying to demonstrate an effect in a very small proportion of our patients."
>
> —Partner organization

Rango could improve participants' treatment adherence only if participants were not already adherent when they enrolled. Data collected by VillageCare indicate that only 17 percent of participants were nonadherent at enrollment. Moreover, interviewees commented that participants who were most likely to actively engage in Rango were those who were already taking good care of themselves, while those who most needed support were less likely to enroll in the first place and less likely to make active use of it if they did. It may be that Rango will have an impact only on the relatively small proportion of participants who were nonadherent at enrollment and who were actively engaged in the program.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of VillageCare's Rango program and the baseline characteristics of the treatment and control groups.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability as of September 1, 2017: VillageCare

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	418[a]
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018	5,388[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	391
Likelihood of all-cause hospitalizations	2,599
MDE sample size requirement to detect 20% effect	
Total expenditures	98
Likelihood of all-cause hospitalizations	650
Participation/Selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	No serious issues. Selected comparison group.
Do claims identify the primary expected effects?	Some effects observed in claims data but important effects likely missing

Table III.1 *(continued)*

Evaluability domain	Response
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	Services utilized

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We are planning to conduct a rigorous impact analysis of the Rango program. We will construct a valid comparison group for Medicare beneficiaries by using propensity score matching to select comparison group members who live in New York City; have HIV/AIDS; and have similar diagnoses, sociodemographic characteristics, and service use as program participants. We plan to construct a similar comparison group for Medicaid beneficiaries when data are available. Our final projected analysis sample (shown in Table III.1) is based on the number of treatment group beneficiaries when enrollment ended in February 2017. This sample should be large enough to detect plausible effects on most claims-based measures for the Medicaid analysis sample and for some claims-based variables for the Medicare analysis sample.

B. Characteristics of Medicare participants at baseline

This section includes a summary of both core and awardee-specific claims-based outcomes at baseline. For the purpose of our evaluation, the treatment group consists of beneficiaries who were in Medicare fee-for-service (FFS) and who were enrolled in Rango (according to lists from the awardee) between April 1, 2015, and November 30, 2016. Medicare managed care, Medicaid FFS, and Medicaid managed care beneficiaries were all excluded from the analysis due to lack of data availability.

VillageCare began to enroll Medicare and Medicaid beneficiaries into the Rango program in April 2015. As of the end of November 2016, the program had 3,625 participants, 869 of whom were enrolled in Medicare (FFS and managed care).[4]

In presenting the baseline characteristics for Medicare FFS beneficiaries, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they had to have been enrolled in the awardee's program on or before November 30, 2016, in order to

[4] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

ensure a sufficient run-out period to capture nearly all claims for the most recent participants. The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. After we excluded patients who did not meet the above criteria, 367 Medicare FFS participants (out of 869 total Medicare participants, FFS and managed care, enrolled through November 30, 2016) were included in the analysis of baseline characteristics for this report. The vast majority of the Medicare beneficiaries excluded from the analysis were in Medicare managed care during their month of enrollment.

The Medicare FFS beneficiaries participating in the Rango program are fairly typical of HIV/AIDS patients nationwide[5] in terms of demographics and health status characteristics (Table III.2). The most common characteristics of participants include being younger than 65 (86 percent); male (66 percent); black (58 percent); originally entitled to Medicare through disability (90 percent, relative to a national average of 24 percent); and dually eligible (65 percent, relative to a national average of 18 percent). Less commonly, 4 percent of beneficiaries have end-stage renal disease (ESRD), according to the original reason of Medicare entitlement. None of the participants are enrolled in hospice. Their mean hierarchical condition category (HCC) risk score is 1.66, relative to a national mean risk score of 1.14 in calendar year 2014 among beneficiaries younger than 65.[6] Their median risk score is 1.38, with a 25th percentile risk score of 1.03 and a 75th percentile risk score of 2.00. Taken together, the scores indicate that the participants are substantially sicker than the average Medicare FFS beneficiary.

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in VillageCare's program through November 30, 2016

	All participants (N = 367)	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
Younger than 65	317	86
65 to 74	47	13
75 to 84	3	0.82
85 and older	0	0
Gender		
Female	123	34
Male	244	66
Race		
White	77	21
Black	214	58

[5] As in our sample, black and Hispanic populations nationwide include high proportions of people with HIV. See http://www.cdc.gov/hiv/statistics/overview/ataglance.html. In addition, most HIV/AIDS patients nationwide are younger than 55. See http://www.cdc.gov/hiv/group/age/olderamericans/index.html.

[6] See https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html.

Table III.2 *(continued)*

Characteristics	All participants (N = 367)	
	Number	Percentage
American Indian, Alaska Native, Asian/Pacific Island American, or other	10	3
Hispanic	66	18
Original reason for Medicare eligibility		
Old age and survivor's insurance	24	7
Disability insurance benefits	330	90
ESRD[a]	13	4
Hospice[b]	0	0
Medicare/Medicaid dual status, percentage dual[b]	240	65
HCC score[c]		Statistic
Mean		1.66
25th percentile		1.03
Median		1.38
75th percentile		2.00

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the beneficiary signed up for the program. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

ESRD = end-stage renal disease; FFS = fee-for-service; HCC = hierarchical condition category.

VillageCare expects the Rango program to reduce total expenditures by 8.6 percent, compared with what they would have been absent the program. The awardee anticipates that the reduction in expenditures will be a result of achieving the program goals, which include (1) improving the engagement and retention of HIV/AIDS patients in care; (2) enabling the development of more timely, tailored interventions to help participants adhere to their HIV medication; (3) increasing participants' time in first-line (that is, least burdensome and least costly) HIV/AIDS treatment; and (4) reducing costly inpatient and outpatient services associated with treatment failure.

Consistent with the high HCC risk scores, participants also have high baseline expenditures and service use (Table III.3). We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments, in total and by major types of services.[7]

[7] All national data in this paragraph are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

There was an upward trend in quarterly expenditures. The total average PBPM Medicare payment during the baseline year was $1,593, relative to a national average of $855 in 2014 for those younger than 65. The average PBPM Medicare payment was $513 for acute inpatient services (relative to a national average of $322), $385 for outpatient services (relative to a national average of $126), and $604 for physician services.

The service use was high for participants relative to the corresponding nationwide rates for the Medicare population younger than 65. The rate of acute care hospitalizations was 361 per 1,000 Medicare FFS beneficiaries per year, relative to a national average among those younger than 65 of 324 per 1,000 beneficiaries per year during the baseline year. Twenty percent of the beneficiaries had at least one hospitalization in the baseline year.[8] The rate of ED visits was 879 per 1,000 Medicare FFS beneficiaries per year, higher than the national average in 2014 of 789 per 1,000 beneficiaries per year. Forty-three percent of participants had at least one ED visit in the baseline year. At baseline, the rate of primary care visits in any setting was 3,410 per 1,000 Medicare FFS beneficiaries per year; 83 percent occurred in an ambulatory setting. The rate of specialty care visits in any setting was 12,150 per 1,000 Medicare FFS beneficiaries per year; 88 percent occurred in an ambulatory setting.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in VillageCare's program through November 30, 2016

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees	367	351	359	367	367
Average Medicare expenditures PBPM[a]					
Total	1,593 (314)	1,422 (305)	1,422 (368)	1,666 (329)	1,838 (368)
Acute inpatient	513 (76)	415 (111)	381 (122)	618 (139)	617 (129)
Inpatient other[b]	27 (12)	49 (36)	0 (0)	22 (14)	36 (23)
Outpatient[c]	385 (43)	311 (35)	369 (45)	399 (60)	454 (60)
Physician services	604 (295)	541 (259)	634 (339)	566 (271)	672 (330)
Home health	27 (6)	28 (9)	16 (7)	24 (9)	40 (11)

[8] All national data in this paragraph except for ambulatory observation stays are from the Centers for Medicare & Medicaid Services, "Public Use File; New Data on Geographic Variation." Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. Accessed February 2016.

Table III.3 *(continued)*

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Skilled nursing facility	18 (15)	33 (25)	13 (13)	24 (31)	5 (5)
Hospice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Durable medical equipment	20 (8)	45 (35)	9 (2)	13 (3)	13 (4)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions[d]	361 (54)	292 (86)	307 (84)	422 (88)	406 (74)
Outpatient ED visits[e]	879 (107)	912 (156)	863 (148)	821 (137)	922 (132)
Primary care visits in any setting	3,410 (286)	2,969 (307)	3,157 (439)	3,584 (349)	3,896 (341)
Primary care visits in ambulatory settings	2,841 (235)	2,502 (265)	2,589 (258)	2,918 (298)	3,326 (305)
Specialist visits in any setting	12,150 (695)	11,129 (898)	11,550 (838)	12,615 (865)	13,225 (849)
Specialist visits in ambulatory settings	10,696 (639)	9,551 (804)	10,460 (761)	10,729 (726)	11,963 (794)
Measures of any health care utilization					
Percentage with a hospital admission[d]	20 (2)	6 (1)	6 (1)	7 (1)	9 (1)
Percentage with an outpatient ED visit[e]	43 (3)	17 (2)	16 (2)	14 (2)	16 (2)
Percentage with a 30-day readmission among all discharges	11 (3)	3 (3)	5 (5)	23 (7)	8 (5)
Percentage of participants with a readmission among all participants	3 (1)	0 (0)	0 (0)	2 (1)	1 (< 0.5)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

Table III.3 *(continued)*

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

[e]Includes visits to an ED, as well as observation stays.

ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

The trend in utilization over time is mixed. Acute hospital admissions rose over the first three quarters, then fell slightly in quarter 4. On the other hand, there was an overall upward trend in both primary care visits and specialist visits in any setting and in ambulatory settings from quarter 1 to quarter 4. In contrast, ED visits declined in quarters 2 and 3, then rose in quarter 4. Overall, participants had higher expenditures and a higher rate of acute care hospitalizations relative to the national averages for Medicare FFS beneficiaries younger than 65. These findings indicate that there may be the potential for improving the care of participating beneficiaries.

People with HIV/AIDS generally have high rates of mental health and substance abuse problems relative to the general population. In the extreme, these problems may lead to an ED visit or a hospital admission. Table III.4 presents two measures specific to VillageCare that relate to mental health and substance abuse. For VillageCare enrollees, the rate of ED visits for mental health or substance abuse is 37 per 1,000 Medicare FFS beneficiaries per year, while the rate of hospital admissions for mental health or substance abuse is 42 per 1,000 Medicare FFS beneficiaries per year.

Table III.4. Measures specific to the awardee for Medicare FFS beneficiaries enrolled in VillageCare's program through November 30, 2016

		Utilization for each quarter in the 12 months before enrollment			
Health care utilization rates (annualized per 1,000)	12 months before enrollment	Q1 (10 to 12 months before enrollment)	Q2 (7 to 9 months before enrollment)	Q3 (4 to 6 months before enrollment)	Q4 (1 to 3 months before enrollment)
Total number of enrollees	367	351	359	367	367
ED visits for mental health or substance abuse—primary diagnosis[a]	37 (11)	35 (20)	11 (11)	44 (23)	55 (24)
Hospital admissions for mental health or substance abuse—primary diagnosis, all hospitals[a,b]	42 (16)	47 (46)	23 (16)	55 (29)	44 (35)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of November 30, 2016.

Note: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as the 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

Table III.4 *(continued)*

[a]The search for mental health and substance abuse diagnoses is limited to the primary diagnosis on the claim.

[b]Unlike the acute hospital admissions measure in Table III.3, the measure for hospital admissions for mental health or substance abuse includes rehabilitation, long-term care, and psychiatric hospitals.

ED = emergency department.

C. Characteristics of Medicaid participants at baseline

In presenting the baseline characteristics for Medicaid beneficiaries, we included both Medicaid FFS and Medicaid managed care beneficiaries. Similar to the restrictions imposed on the Medicare FFS beneficiaries above, we restricted the treatment group to Medicaid beneficiaries who were enrolled in Medicaid when their eligibility for awardee-provided services began (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). Due to delays in Medicaid availability, we can only include beneficiaries who were enrolled in the awardee's program on or before June 30, 2015. After we excluded participants who did not meet the above criteria, 530 Medicaid participants (out of 3,000 total Medicaid participants enrolled through May 31, 2016) were included in the analysis of baseline characteristics for this report. The vast majority of the Medicaid participants excluded from the analysis were those who enrolled after June 30, 2015.

All of the Medicaid beneficiaries included in this analysis were enrolled in Rango in the second quarter of 2015. The most common characteristics of Medicaid participants (Table III.5) include being 45 to 54 years old (46 percent), male (69 percent), black (62 percent), eligible for full Medicaid benefits (98 percent), in the eligibility category of supplemental security income (SSI) for blind/disabled (51 percent), enrolled in a comprehensive managed care plan (72 percent), and non-dual eligible (76 percent). In addition, most Medicaid participants were not enrolled through a home and community-based services (HCBS) waiver (98 percent) and most had no third-party insurance (98 percent).

In Table III.6, we find that the most common Chronic Disability Payment System (CDPS) categories of Medicaid participants include infectious disease (86 percent), psychiatric disorders (49 percent), cardiovascular (43 percent), substance abuse (35 percent), and pulmonary (33 percent). Ninety-five percent of participants are in at least one CDPS category. Although only 86 percent of participants had an infectious disease diagnosis in their claims data in the 365-day baseline period, it is anticipated that all participants will have a diagnosis of HIV/AIDS (an infectious disease) in their claims data at some point in time (that is, either in the baseline period or before the baseline period), because HIV/AIDS is a requirement to qualify for Rango. The high rates of substance abuse (35 percent) and psychiatric problems (49 percent) are consistent with the high rates for these conditions that are found in the HIV/AIDS population nationwide. Participants have a mean risk score of 3.17, a median risk score of 2.91, a 25th percentile risk score of 2.18, and a 75th percentile risk score of 3.94.

Table III.5. Baseline year demographic characteristics of Medicaid beneficiaries enrolled in VillageCare's program through the first program quarter (June 30, 2015)

	All enrollees (N = 530)[a]	
Characteristics	**Number**	**Percentage**
Age as of enrollment date		
22-34	40	7.55
35-44	79	14.91
45-54	245	46.23
55-64	140	26.42
65-74	26	4.91
Gender		
Female	166	31.32
Male	364	68.68
Race and ethnicity		
White	44	8.30
Black	329	62.08
Asian or Pacific Islander	2	0.38
American Indian, Alaska Native, or other	2	0.38
Hispanic	1	0.19
Hispanic and one or more races	134	25.28
More than one race (not Hispanic)	5	0.94
Type of benefits		
Full Medicaid benefits	517	97.55
Restricted benefits	13	2.45
Medicaid eligibility category		
SSI aged	8	1.51
Non-SSI aged	2	0.38
SSI blind/disabled	268	50.57
Non-SSI blind/disabled	37	6.98
TANF, safety net, or low-income family adults	205	38.68
All other adults	10	1.89
Managed care enrollment		
Comprehensive managed care plan	380	71.70
Long-term care carve out	28	5.28
No managed care enrollment	122	23.02
Medicare/Medicaid dual status, percent dual		
Dual	129	24.34
Non-dual	401	75.66
HCBS waiver enrollment		
Enrolled in any HCBS waiver	9	1.70
Not enrolled in a HCBS waiver	521	98.30

Table III.5 *(continued)*

Characteristics	All enrollees (N = 530)[a]	
	Number	Percentage
Third-party insurance		
Third-party insurance	11	2.08
No third-party insurance	519	97.93
Quarter of initial program enrollment		
Q2 2015	530	100.00

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of June 30, 2015.

Note: The baseline period is defined as one year (365 days) prior to each individual beneficiary's program enrollment date. Medicaid enrollees must have at least 90 days eligibility in the baseline period and on the date of program enrollment to be included in the eligible sample. All beneficiary characteristics (other than CDPS category and risk score) are measured in the last month of the baseline period.

[a]Excludes Medicaid beneficiaries who have state plan enrollment during the month of program enrollment.

HCBS = Home and Community Based Services; SSI = Supplemental Security Income; TANF = Temporary Assistance for Needy Families.

Table III.6. CDPS categories of Medicaid beneficiaries enrolled in VillageCare's program through the first program quarter (June 30, 2015)

Characteristics	All enrollees (N = 530)[a]	
	Number	Percentage
Selected CDPS categories[b]		
Beneficiaries in one or more CDPS categories	506	95.47
Infectious disease	455	85.85
Psychiatric	261	49.25
Cardiovascular	229	43.21
Substance abuse	184	34.72
Pulmonary	173	32.64
Beneficiaries not in a CDPS category	24	4.53
CDPS risk score[b]		
Mean	3.17	
25th percentile	2.18	
Median	2.91	
75th percentile	3.94	

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of June 30, 2015.

[a]Excludes Medicaid beneficiaries who have state plan enrollment during the month of program enrollment.

[b]Categories and risk scores are defined using the CDPS software and are based on the ICD-9 codes on Medicaid claims occurring during each beneficiary's one-year baseline period. The five most common conditions among beneficiaries in the sample are reported in this table.

CDPS = Chronic Disability Payment System.

Baseline expenditure and utilization statistics for Medicaid participants are presented in Tables III.7 and III.8 for non-dual status participants (enrolled in Medicaid only) and for dual status participants (enrolled in Medicare and Medicaid), respectively. Unlike the beneficiary characteristics presented in Tables III.5 and III.6, the expenditure and utilization analysis presented in Tables III.7 to III.9 excludes Medicaid beneficiaries who have partial benefits or third-party benefits during the month of program enrollment (as well as those who have state plan enrollment).

In Table III.7, we find that the total average PBPM Medicaid payment among Medicaid non-dual status participants during the baseline year was $5,445. The average PBPM Medicaid payments for non-dual status participants were $368 for acute inpatient stays; $26 for ED visits; $3,527 for pharmacy services; and $1,524 for other services. The large proportion of total spending accounted for by pharmacy expenditures is consistent with the high cost of the drugs commonly used to treat HIV/AIDS. Overall, total expenditures fell in both the second and third quarters before rising back to their quarter 1 values in quarter 4. In contrast, the expenditure components displayed varying trends. For example, "other" expenditures fell each quarter, whereas pharmacy expenditures fell from quarter 1 to quarter 2 and rose in quarter 3 and quarter 4. Medicare and Medicaid cover different services, making it difficult to compare the Medicare expenditures found in Table III.3 with the Medicaid expenditures found in Table III.7. For example, New York State Medicaid covers prescription drug spending, whereas Medicare (Parts A and B) does not cover prescription drug spending.[9]

Table III.7 also provides utilization data for the Medicaid non-dual status participants. The rate of acute hospital admissions was 517 per 1,000 Medicaid beneficiaries per year. The rate of ED visits was 1,314 per 1,000 Medicaid beneficiaries per year, with 15 percent (198 of 1,314) leading to an inpatient stay. Acute hospital admissions and ED visits exhibited no clear trend over time.

[9] Medicare Part D does cover prescription drugs. However, due to a lack of data availability, Part D drug spending was not included in the analysis. Furthermore, many Medicare beneficiaries are not enrolled in Medicare Part D.

Table III.7. Baseline year expenditures and health care utilization for Medicaid non-dual status beneficiaries enrolled in VillageCare's program through the first program quarter (June 30, 2015)

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Quarter 1 (10 to 12 months before enrollment)	Quarter 2 (7 to 9 months before enrollment)	Quarter 3 (4 to 6 months before enrollment)	Quarter 4 (1 to 3 months before enrollment)
Total number of enrollees[a]	393	375	380	390	393
Average Medicaid expenditures PBPM[b]					
Total payment	5,445 (202)	5,606 (289)	5,339 (252)	5,134 (278)	5,686 (296)
Acute inpatient stays	368 (62)	432 (119)	473 (107)	277 (69)	295 (88)
Total ED payment	26 (3)	28 (4)	28 (5)	23 (4)	25 (4)
ED visits that lead to an inpatient stay	2 (<0.5)	3 (1)	2 (1)	1 (1)	2 (1)
ED visits that do not lead to an inpatient stay	24 (3)	25 (4)	27 (5)	21 (4)	23 (3)
Pharmacy	3,527 (141)	3,528 (216)	3,278 (162)	3,349 (194)	3,926 (263)
Other[c]	1,524 (79)	1,618 (88)	1,560 (87)	1,485 (121)	1,441 (77)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions	517 (74)	567 (123)	525 (105)	542 (108)	439 (90)
Total ED visits	1,314 (117)	1,427 (245)	1,446 (245)	1,136 (144)	1,254 (172)
ED visits that lead to an inpatient stay	198 (41)	316 (93)	193 (61)	135 (49)	153 (60)
ED visits that don't lead to an inpatient stay	1,116 (97)	1,111 (158)	1,253 (207)	1,001 (127)	1,101 (139)

Source: Mathematica analysis of information from awardee's finder file and Medicaid claims and enrollment data as of June 30, 2015.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as the 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter is 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and each baseline quarter according to the number of days that each beneficiary is enrolled in Medicaid during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Excludes Medicaid beneficiaries who have state plan enrollment, partial benefits, or third party benefits during the month of program enrollment.

Table III.7 *(continued)*

[b]Total Medicaid expenditures for the baseline year or a given quarter exclude capitated payments. Expenditures were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[c]Other services include non-office visit practitioner services, dental care, eye care, home health care, laboratory services, intermediate care facility services, nursing home services, child care, and clinic services.

ED = emergency department; PBPM = per beneficiary per month.

In Table III.8, we find that the total average PBPM Medicaid payment among Medicaid dual status participants during the baseline year was $1,589. The average PBPM Medicaid payments for dually eligible participants were $43 for acute inpatient stays; $4 for ED visits; $240 for pharmacy services; and $1,302 for other services. Common other services received by participants include home health, laboratory, intermediate care facility, and clinic services. Spending on these service types will be broken out separately in future reports. For dually eligible beneficiaries, Medicare is typically the primary payer. Thus, the majority of spending for dual eligibles will be paid for by Medicare. In contrast, for non-dual status Medicaid beneficiaries, Medicaid is typically the primary payer. Thus, Medicaid pays for the majority of the services for non-dual status beneficiaries. As a result, the total dual status Medicaid spending ($1,589) is quite low relative to the total non-dual status Medicaid spending ($5,445). There is a clear downward trend in total expenditures over time, which is driven partly by the downward trend over time in pharmacy expenditures and partly by the downward trend in "other" expenditures in the last two quarters.

Table III.8 also provides utilization data for the Medicaid dually eligible participants. The rate of acute hospital admissions was 434 per 1,000 Medicaid beneficiaries per year. The rate of ED visits was 1,127 per 1,000 Medicaid beneficiaries per year, with 12 percent (139 of 1,127) leading to an inpatient stay. There was no clear trend in service use over the baseline period.

Table III.8. Baseline year expenditures and health care utilization for Medicaid dually eligible beneficiaries enrolled in VillageCare's program through the first program quarter (June 30, 2015)

Types of expenditures and utilization measures	12 months before enrollment	Expenditures and utilization for each quarter in the 12 months before enrollment			
		Q1 (10 to 12 months before enrollment)	Q2 (7 to 9 months before enrollment)	Q3 (4 to 6 months before enrollment)	Q4 (1 to 3 months before enrollment)
Total number of enrollees[a]	114	106	107	112	114
Average Medicaid expenditures PBPM[b]					
Total Payment	1,589 (203)	1,808 (301)	1,690 (258)	1,537 (200)	1,244 (144)
Acute inpatient stays	43 (23)	32 (11)	34 (12)	51 (14)	50 (14)
Total ED Payment	4 (1)	3 (1)	5 (2)	3 (1)	4 (2)
ED visits that lead to an inpatient stay	0 (<0.5)	0 (<0.5)	0 (<0.5)	0 (<0.5)	0 (<0.5)
ED visits that don't lead to an inpatient stay	4 (1)	3 (1)	5 (2)	2 (1)	4 (2)
Pharmacy	240 (76)	519 (215)	195 (78)	174 (70)	89 (49)
Other[c]	1,302 (167)	1,255 (154)	1,456 (228)	1,310 (181)	1,101 (130)
Health care utilization rates (annualized per 1,000)					
Acute hospital admissions	434 (228)	345 (109)	337 (120)	476 (133)	534 (146)
Total ED visits	1,127 (192)	1,417 (755)	1,496 (358)	805 (216)	819 (195)
ED visits that lead to an inpatient stay	139 (43)	115 (65)	112 (64)	220 (86)	107 (61)
ED visits that don't lead to an inpatient stay	989 (171)	1,302 (749)	1,384 (319)	586 (168)	712 (189)

Source: Mathematica analysis of information from awardee's finder file and Medicaid claims and enrollment data as of June 30, 2015.

Notes: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as a 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter is 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and in each baseline quarter according to the number of days that each beneficiary was enrolled in Medicaid during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Excludes Medicaid beneficiaries who have state plan enrollment, partial benefits, or third party benefits during the month of program enrollment

[b]Total Medicaid expenditures for the baseline year or for a given quarter excludes capitated payments and were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[c]Other services include non-office visit practitioner services, dental care, eye care, home health care, laboratory services, intermediate care facility services, nursing home services, child care, and clinic services.

ED = emergency department; PBPM = per beneficiary per month.

As mentioned, people with HIV/AIDS generally have high rates of mental health and substance abuse problems relative to the general population. In the extreme, mental health and substance abuse problems may lead to an ED visit or a hospital admission. Table III.9 presents two measures specific to VillageCare that relate to mental health and substance abuse. The rate of ED visits for mental health or substance abuse is 101 per 1,000 Medicaid beneficiaries per year, while the rate of hospital admissions for mental health or substance abuse is 125 per 1,000 Medicaid beneficiaries per year.

Table III.9. Measures specific to VillageCare for Medicaid beneficiaries enrolled in the program through the first program quarter (June 30, 2015)

Health care utilization rates (annualized per 1,000)	12 months before enrollment	Utilization for each quarter in the 12 months before enrollment			
		Q1 (10 to 12 months before enrollment)	Q2 (7 to 9 months before enrollment)	Q3 (4 to 6 months before enrollment)	Q4 (1 to 3 months before enrollment)
Total number of enrollees[a]	507	481	487	502	507
ED visits for mental health or substance abuse—primary diagnosis[b]	101 (23)	93 (30)	117 (42)	97 (36)	95 (33)
Hospital admissions for mental health or substance abuse—primary diagnosis[b]	125 (30)	119 (40)	117 (46)	154 (56)	111 (37)

Source: Mathematica analysis of information from awardee's finder file and Medicaid claims and enrollment data as of June 30, 2015.

Note: The baseline year is the 365 days before each participant's enrollment date. Each baseline quarter is defined as the 91-day period starting from each beneficiary's enrollment date. For example, the fourth baseline quarter is the 91 days before each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, and so on. The first baseline quarter will be 92 days, thereby summing to 365 days. We weighted every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in Medicaid during the baseline year and in each baseline quarter.

Standard errors are shown in parentheses.

[a]Excludes Medicaid beneficiaries who have state plan enrollment, partial benefits, or third party benefits during the month of program enrollment.

[b]The search for mental health and substance abuse diagnoses is limited to the primary diagnosis on the claim.

ED = emergency department.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

VillageCare was still developing its payment model at the end of the third year of implementation. The awardee planned to continue discussions after the no-cost extension with the MCOs that provided patient referrals for this HCIA R2 demonstration. However, VillageCare was not actively negotiating for payment at the end of the cooperative agreement.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: (1) the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, (2) narratives and payment model reports the awardee submitted to the implementation and monitoring contractor, and (3) relevant information that arose during our third round of interviews with the awardee. We did not interview the awardee about its payment model because it did not advance from the development stage to active negotiations with one or more payers.

B. Description of the payment model

The payment model developed by VillageCare consists of a PBPM payment to the program from an entity that is responsible for the health of its population, such as an accountable care organization or MCO. The awardee leaders described two levels of payment, one intended to cover just the Rango software and another higher level to cover the software plus VillageCare staff who run the program, such as the enrollment staff and health coaches.

C. Status of the payment model

VillageCare was still in a development phase for its payment model at the end of the third year of implementation. The awardee planned to continue its discussions with the MCOs that provided patient referrals, but it was not in active negotiations for payment. To make the case to potential payers that Rango was worth the PBPM payment, VillageCare needed evidence of Rango's effectiveness. The awardee had agreements with three evaluators and had received preliminary studies from two of them. It was expecting further results in fall 2017. The evaluators were looking at the impact of VillageCare's intervention on costs and outcomes.

D. Factors associated with the development of the payment model

The primary barrier to the development and implementation of a payment model for this awardee was the preference of potential payers to have evaluation results prior to any payment negotiations. VillageCare stated that potential clients were holding off on any contracting decisions until they saw evidence from the evaluations of Rango's impact on participants' health outcomes.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

In an effort to sustain the Rango program, VillageCare worked on gaining approval from its two MCO partners for the payment model. One MCO was interested in providing payment and expanding the program to its other members, if Rango showed evidence of effectiveness and the price was low enough. The awardee also was trying to obtain grants from federal agencies and philanthropic foundations. In addition, VillageCare was considering scaling the program to other providers and populations. If the awardee could not secure funding during the no-cost extension, its leaders said they would have to discontinue the program.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, VillageCare began actively pursuing strategies to sustain its program, including setting a PBPM fee for program services that it marketed to its partners. The awardee also began analyzing clinical outcomes data to demonstrate the value of the Rango intervention to potential payers. One of the awardee's MCO partners had said that the PBPM for all services was too expensive; the MCO requested a lower PBPM to cover just the technology component. The awardee had no plans to scale or replicate the program at that time.

C. Implementing the SSR plan: Progress and changes

Sustainability. In Year 3, the awardee hoped to have funding in place by the end of the no-cost extension to sustain the program without interruption. VillageCare was exploring whether its partner organizations would be willing to pay to continue using Rango. The awardee was also pursuing funding from private foundations.

VillageCare developed two possible PBPM payments: one that covered the full services of Rango and one that only covered the technology. The awardee was discussing with its partner organizations whether they would be willing to pay the PBPM to continue offering Rango to their patients. The community-based organizations and primary care providers that VillageCare partnered with were unable to pay for continued use of Rango because they had no means to recover the costs. However, the MCOs might pay for Rango if the awardee could provide evidence that the program improved outcomes and therefore might reduce costs. The awardee was still waiting for results from the evaluations it had funded to present to the MCOs. While waiting for the results, VillageCare remained in discussion with the MCOs, but it was not yet close to having contracts with either of them.

The awardee planned to seek foundation funding to help sustain the program in the short term while it sought contracts with other providers or MCOs for higher levels of ongoing funding. VillageCare had identified three potential foundations (the Elton John AIDS Foundation, the Robin Hood Foundation, and the New York Community Trust) whose proposal and funding requirements aligned with the Rango program. The awardee had one such proposal in process and was hopeful that funding could be available within a few months. The awardee also attempted to obtain funding from the New York City Council but was unsuccessful; it planned to try again in 2018.

Finally, the awardee was in discussion with its primary care provider partners about teaming up to seek funding through National Institutes of Health (NIH) or other grants. VillageCare felt that partnering with the primary care providers could significantly strengthen the chances of obtaining such funding because of the primary care providers' experience in obtaining grants. Researchers at Mount Sinai expressed particular enthusiasm about partnering with the awardee to pursue funding.

Scalability. If it could justify a business case, one MCO partner was interested in potentially adopting the Rango tool for all of its current members. The MCO looked at Rango as a way to gain new business—essentially marketing it as a benefit of enrollment that would be exclusive to the health plan. This would be a collaboration with VillageCare in which the MCO would purchase and manage the day-to-day use of the tool, while VillageCare would retain the data analysis and content development tasks. The MCO would expect to make some enhancements, including broadening the features so that the program could serve as the corporate platform for the MCO's HIV-positive members and modifying the tool for its HIV-negative transgender population.

VillageCare thought the program could also be applied to health care issues other than HIV and potentially even issues beyond health care. In fact, it has been contacted by an outreach project for sex workers and individuals working on housing stability issues. However, a

consultant advised the awardee that it would be best to initially stay focused on the HIV-positive population because that is the population for which it could have data demonstrating the program's success.

Replicability. There were no signs that the Rango program had been replicated by other organizations, nor did the awardee appear active in efforts to help others replicate it.

D. Factors associated with progress toward implementing the SSR plan

Lack of outcome data to demonstrate that Rango is a cost-effective program inhibited the awardee's ability to secure contracts with MCOs and other entities for ongoing funding, and therefore presented the main challenge to SSR.

In addition, the program director noted that the decision to outsource the platform was made early on, as app development seemed too far from VillageCare's capabilities. However, this decision left the awardee reliant on its technology partners for Rango's future.

If the awardee did not have funding in place by the end of the cooperative agreement, program leaders thought continuing the program would be very difficult. The awardee could reinstate the program later, although there were risks to long gaps in service. Encouraging former users of the program to engage anew could be challenging, and VillageCare would likely lose some participants. One respondent noted that the program's ability to foster relationships in real time through information sharing was what kept members engaged: "If you put that on ice for a certain number of months, you lose the stickiness and people might not come back."

Further, the awardee found that forming relationships with new organizations to scale the program was difficult, particularly for organizations that focus on different patient populations. Even reaching out to new organizations in the same field involves a significant business development effort that the awardee doubted it has the resources or expertise to undertake.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the awardee received a no-cost extension, it will not enroll new beneficiaries in Rango. Therefore, the implementation evaluation team has concluded primary data collection, unless CMS requests a follow-up with the awardee to discuss the status of its payment model or SSR plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

The next step in the impact evaluation is to conduct propensity score matching for both Medicaid and Medicare beneficiaries for all beneficiaries who enrolled through February 28, 2017. After our initial attempts to conduct propensity score matching for Medicare beneficiaries who enrolled during the first two years, we identified unexpected issues about the services participants received on their date of enrollment that have important implications for the selection of comparison groups. We have developed strategies to deal with this issue, looking at service use on the day of enrollment and during the following three months for the treatment group to ensure that the comparison group we select is well matched on these measures. We are now working to complete the selection of a final comparison group for Medicare beneficiaries. We will use similar strategies to select a comparison group for Medicaid beneficiaries as soon as Medicaid data are available.

After we select Medicare and Medicaid comparison groups, we will conduct impact analyses that compare outcomes for the program participants to their matched comparison counterparts. We will also perform extensive sensitivity and robustness checks on all results. In addition, we will consider expanding our analyses of outcomes to potentially include a measure for whether the beneficiary filled a prescription for HIV/AIDS (a proxy for drug treatment adherence) and for the all-cause mortality rate. Finally, we will estimate program impacts in a Bayesian framework, which will permit the estimation of each program impact to borrow strength from estimates from other time periods and outcomes and also allow us to estimate the probability that a program impact exceeds a specified value.

We will provide CMS with final impact results for VillageCare's Medicare and Medicaid participants as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE,WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Washington University School of Medicine in St. Louis

April 18, 2018

Margaret Gerteis (Mathematica Policy Research)
Stefanie Pietras (Mathematica Policy Research)
Lori Timmins (Mathematica Policy Research)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

The Washington University School of Medicine used funding from HCIA R2 to create the Contraceptive Choice Center (C3), an innovation designed to address the problem of unintended

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Washington University School of Medicine received a three-month extension through November 30, 2017, to continue to provide services.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

pregnancy and childbirth among high-risk women in the St. Louis, Missouri, metropolitan area by improving women's access to effective methods of contraception (Table I.1). The C3 program was modeled after the earlier Contraceptive CHOICE Project (which ran from 2007 to 2011), a project supported entirely by private funding that combined comprehensive contraceptive counseling and support with same-day access to long-acting reversible contraceptives (LARCs) at no cost to patients.[4] Through its HCIA R2 cooperative agreement, Washington University set out to (1) determine whether the model could be effectively implemented in the absence of private financing by using Medicaid, commercial insurance, and Title X family planning funding; (2) demonstrate the cost-effectiveness of the model to payers; and (3) develop a payment model that would reduce barriers to access by incentivizing providers to adopt the C3 approach. The three major components of the C3 service delivery model included (1) employing and training a social worker to serve as a federally certified and state-licensed insurance navigator to help patients review their insurance options and apply for coverage (patient navigation); (2) using trained, non-clinician health educators to give structured, evidence-based contraceptive counseling to all patients before they received services and support to those who had concerns after they received services (patient engagement); and (3) having trained clinicians provide same-day contraceptive services (including same-day insertion of LARCs when possible) that followed evidence-based guidelines (direct care provision). The awardee also used HCIA R2 funds to conduct a research study of C3 that was approved by the university's institutional review board (IRB).

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	To provide evidence-based contraceptive counseling and family planning services, including same-day insertion of LARCs (for example, intrauterine devices and implants) to reproductive-age women in St. Louis, MO
Major innovation	Implementing the CHOICE model of comprehensive contraceptive care with Medicaid, commercial insurance, and Title X family planning funding; demonstrating the cost-effectiveness of the model to payers; developing a payment model that reduces barriers to access
Program components	• **Patient navigation** to help patients review insurance options and apply for coverage • **Patient engagement** to give structured, evidence-based contraceptive counseling and support to all patients • **Direct care provision** to provide same-day contraceptive services that follow evidence-based guidelines
Target population	Reproductive-age women 14 years and older in the St. Louis area who are at risk for unintended pregnancy and childbirth
Theory of change/ theory of action	Reducing barriers to evidence-based methods of contraception among women of childbearing age and among clinicians who provide family planning services will increase the uptake of methods proven to be most effective, resulting in a reduction of unintended pregnancies and childbirth and their associated costs.

[4] Secura, G. M., T. Madden, C. McNicholas, J. Mullersman, C. M. Buckel, Q. Zhao, and J. F. Peipert. "Provision of No-Cost, Long-Acting Contraception and Teenage Pregnancy." *New England Journal of Medicine,* vol. 371, 2014, pp. 1316–1323.

Table I.1 *(continued)*

Program characteristic	Description
Payment model	Bundled or episode payment
Award amount	$4,034,879
Effective launch date[a]	January 8, 2015
Program setting	Clinic housed in the Division of Clinical Research, Washington University Department of Obstetrics and Gynecology
Market area	Urban
Market location	St. Louis, MO, metropolitan area
Target outcomes	• Increase in uptake of LARCs to 50% of new contraceptive methods • Increase in contraceptive continuation and satisfaction • Reduce rate of unintended pregnancy by 10% • Reduce costs associated with unintended births by 15%

[a]After the initial planning period, the awardee's program became operational as of this date.

LARC = long-acting reversible contraceptive.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that Washington University was successful in implementing its C3 program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, Washington University enrolled 3,022 direct participants in the C3 program (75 percent of its revised enrollment projection) by the end of the three-year cooperative agreement. Second, the awardee established the C3 clinic and delivered services in a manner consistent with the design proposed in its HCIA R2 application, and successfully adapted the CHOICE model to address financial barriers to access and potential delays in service delivery related to insurance coverage. Third, the awardee implemented flexible staffing models and training to adapt to evolving service delivery needs. Fourth, the awardee's self-monitoring statistics indicated high levels of patient uptake of the most effective evidence-based contraceptive methods and high rates of patient satisfaction with C3 services. Finally, participating clinicians and other implementation staff reported that the program had a positive effect on the delivery of care by providing needed services to women with limited means and limited access to primary or preventive care for whom avoiding unplanned pregnancy was a high priority. Program staff also cited Title X data showing that C3 patients were much more likely to adopt highly effective contraceptive methods than patients in other Title X–funded family planning clinics in Missouri.

Impact evaluation. Due to the lack of a strong comparison group, we do not anticipate being able to conduct a rigorous impact analysis for Washington University.

Payment model. The awardee developed a model that would provide one bundled payment for a 90-day episode of contraceptive care, covering (1) initiation of the contraceptive method (including non-clinician contraceptive counseling, medical intake, clinician services, and pregnancy testing); (2) short-term follow-up and support; (3) an insertion fee modifier for LARCs (if applicable); (4) facility and administrative charges, (including insurance navigation and assistance); and (5) a dispensing fee for LARCs (if applicable). The awardee reported having had positive communications about the proposed payment model with Missouri HealthNet (the

state Medicaid agency) and with a managed care payer. However, changes in state-level government, Medicaid management, and the Missouri Uninsured Women's Health Program slowed Washington University's progress.

Sustainability plans. Washington University plans to sustain the C3 program after the HCIA R2 cooperative agreement ends. C3 leaders are assessing the staffing and budgetary requirements needed to continue the clinic's contraceptive counseling and clinical services, relying on funding streams from Title X and third-party billing for clinical services. Absent other sources of support, they foresee scaling back on evening and weekend hours, research support activities, and possibly on insurance navigation assistance. C3 leaders are also exploring ways to expand the program's scope of work, both as a means of enhancing revenues and in response to identified areas of need (for example, preventive services for the LGBTQ community). They are also interested in possibly extending C3 contraceptive services to underserved populations through the use of mobile units, but recognize that this would require new investments in research and development. If Missouri HealthNet pursues the bundled payment model for contraceptive care that the awardee has proposed, then core features of the C3 model could be replicated and reimbursed in other settings that serve Medicaid patients.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey, which was fielded from July 2016 to October 2016 with 13 contraceptive counselors, medical assistants, administrative staff, and a social worker, achieved a response rate of 100 percent. The clinician survey, which was fielded from March 2017 to June 2017 with 10 physicians and nurse practitioners, also achieved a response rate of 100 percent. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of the participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that Washington University was successful in implementing its C3 program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

To reach its target population of reproductive-age women (14 years and older) at risk for unintended pregnancy and childbirth, Washington University developed and continually refined a four-pronged outreach and recruitment strategy that included the following components: (1) outreach to commercially insured Washington University students and medical center staff and to students at other nearby colleges and universities, (2) partnerships with community organizations serving high-risk populations, (3) digital media outreach, and (4) traditional print and broadcast advertising.

To enroll patients, C3 staff screened new recruits to determine and document their eligibility, based on the HCIA R2 cooperative agreement, Washington University's IRB-approved C3 research study criteria, and Title X requirements and restrictions. Women who had

a hysterectomy or were postmenopausal were not eligible for enrollment in the HCIA R2 demonstration, the C3 research study, or Title X clinic services. To meet Title X requirements, the awardee could not require women to participate in C3 research as a condition of receiving services. Therefore, eligible women could opt out of the C3 research, receive clinical services only, and still be counted as HCIA R2 participants. The clinic's Title X services were not restricted to contraceptive care. However, Title X did not cover and C3 did not provide pregnancy-related care or services. All HCIA R2 and C3 research-eligible women received pregnancy tests at intake. Those who tested positive were screened out and referred elsewhere for clinical services.

For HCIA R2 reporting purposes, Washington University considered as direct participants all women who completed paperwork for the C3 research study or for clinical services only during their first visit. When direct participants returned to C3 for additional clinical services, they were considered indirect participants. That is, some women were counted as both direct and indirect participants.

b. Evidence of enrollment effectiveness

Washington University was partially successful in achieving its enrollment targets. The awardee reported enrolling 3,022 direct participants in its C3 program from January 2015 (when it launched its program) through August 2017, which represents about 75 percent of its 4,047 adjusted three-year projected direct participants (Figure II.1). The awardee's initial target of 10,000 direct enrollees was predicated partly on the assumption that Missouri would expand Medicaid eligibility. However, Missouri did not expand its Medicaid program and the state's policy environment became increasingly unsupportive of women's reproductive health care during the three-year cooperative agreement. The awardee subsequently revised its direct enrollment targets downward to be more in line with its actual enrollment experience. Although quarterly enrollment numbers continued to fall short of the revised targets throughout the cooperative agreement, the awardee estimated that the number of women served would be sufficient for Washington University's C3 research to demonstrate the clinic's effectiveness, given high rates of LARC uptake and anticipated access to Medicaid comparison data in Missouri.

Washington University also reported enrolling 1,548 indirect participants (repeat visitors) from January 2015 to August 2017, which represents about 72 percent of its 2,160 adjusted three-year projected indirect participants (Figure II.2). The awardee's projection for indirect participants was based on the estimation that 40 percent of direct participants would return for additional visits. Therefore, the awardee was below projection in indirect program participants because enrollment of direct participants was below projection. Still, the awardee reported more return appointments per patient than originally anticipated and believed the return visits reflected a high level of patient trust and satisfaction with care at C3.

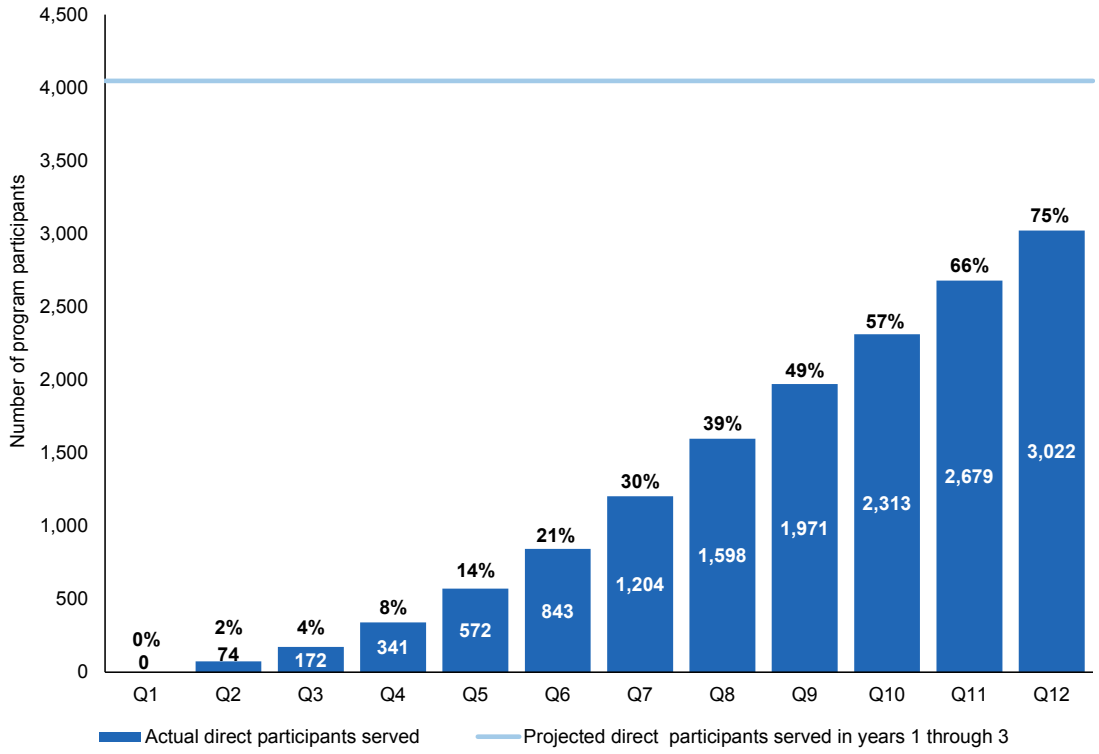
According to Washington University's self-reported enrollment statistics, C3 succeeded in reaching an ethnically diverse patient population that was generally reflective of the population of the St. Louis metropolitan area: 47 percent of enrollees identified as black or African American, 35 percent as white, and 11 percent as Hispanic or Latino. Enrolled patients also spanned the reproductive-age spectrum, including adolescence (8 percent were age 18 or

younger), young adults (39 percent were age 19 to 25), and older adults (53 percent were age 26 or older).

Although nearly half (46 percent) of C3 patients had private insurance, enrollment statistics also indicated that C3 reached lower-income women who faced financial challenges. About 20 percent of participants were covered by Medicaid (or the Medicaid family planning waiver demonstration), but 31 percent of C3 patients were uninsured at their first clinic visit.

The awardee's self-monitoring data provided additional insight into its effectiveness in enrolling women at risk for unintended pregnancy. As of August 2017, self-reports at intake from over 2,300 C3 patients indicated that about 42 percent had previously experienced unintended pregnancies. Although about 28 percent of new patients were already using LARC methods, about 35 percent of new patients reported using either less-effective contraceptive methods or none at all during their last sexual intercourse. Approximately 85 percent of enrolled patients reported wanting to postpone pregnancy for two years or more or to avoid pregnancy altogether.

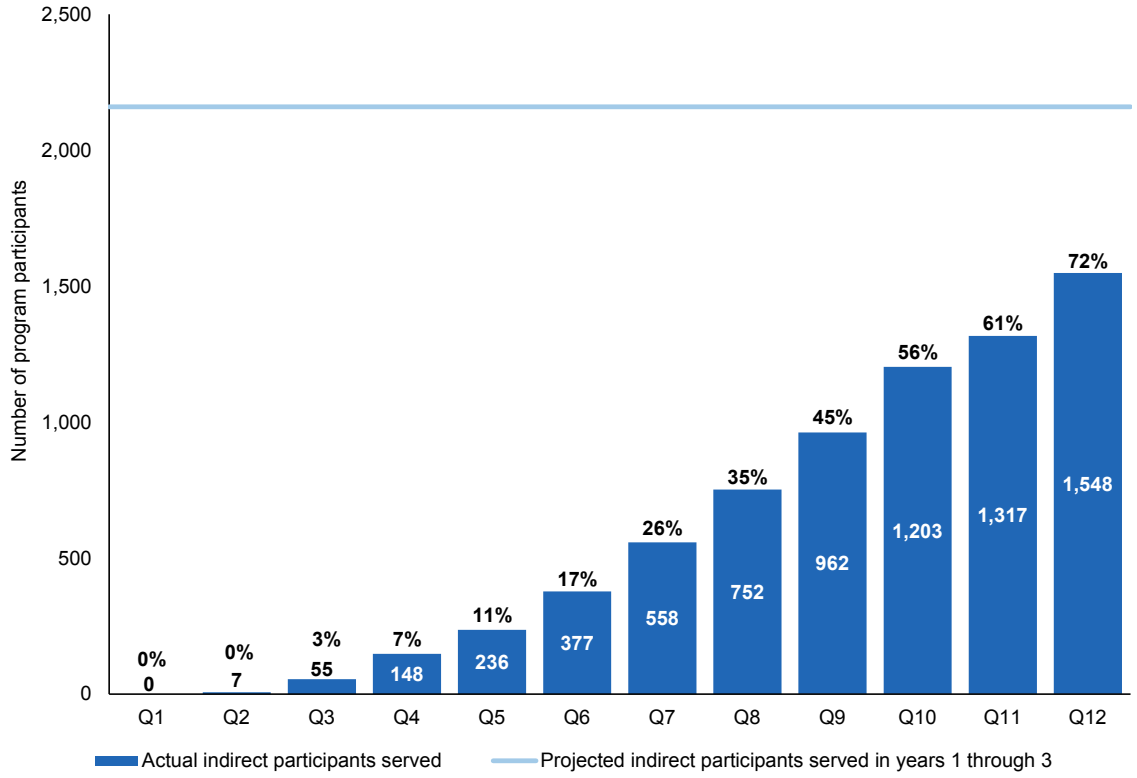
Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. Patients are counted as direct participants at their first visit to the clinic.

Figure II.2. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers. Indirect program participants are repeat visitors to the clinic.

c. Barriers and facilitators associated with enrollment effectiveness

Washington University's progress in meeting its three-year enrollment goals was influenced by several factors. Barriers to enrollment effectiveness included (1) challenges establishing a referral network within the provider community, (2) competing priorities among community organizations serving high-risk populations, (3) lack of awareness about the program within the target population, and (4) a policy environment unsupportive of women's reproductive health care. Facilitators of enrollment effectiveness included (1) outreach strategies and messaging developed with help from marketing consultants and patient and community advisors and (2) word of mouth and referrals within the patient and provider community. Below, we further describe the factors affecting patient referrals to C3 from the provider community, outreach to community organizations, direct outreach to the target population through advertising and digital media, and the influence of the policy and political environment.

First, the awardee emphasized that it took time and continuous effort to establish a reliable referral network among providers. Leaders noted that it was difficult to build awareness of the

program in the provider community, establish a niche for C3 in the landscape of community clinics, and change long-standing habits of providers who typically referred to other established and better-known providers such as Planned Parenthood. By the third year of the HCIA R2 cooperative agreement, however, staff reported that a number of providers in the St. Louis community routinely referred patients to C3, especially for LARC. They estimated that such referrals accounted for about 15 percent to 20 percent of new clinic patients.

Second, throughout the three-year cooperative agreement, C3 staff reported challenges with eliciting referrals from community organizations that serve high-risk populations because the organizations often had other priorities. The awardee reported making a concerted effort to reach out to organizations, such as federally qualified health centers, schools, the county jail, immigrant organizations, and programs for teens, but found that these efforts yielded mixed results. In interviews, staff reported that most community organizations responded favorably to C3's informational outreach about contraceptive care, but they typically did not make referrals to the clinic if pregnancy prevention was not a priority for the organization.

Third, enrollment efforts were hindered by a lack of awareness of the program's services among the target population. To support direct outreach efforts, the C3 staff solicited advice from patients and community advisors during the second year of the HCIA R2 cooperative agreement and engaged marketing and media consultants to help with strategies and messaging. With this guidance, the clinic's staff and leaders worked to establish the C3 brand, emphasizing its fast-track appointments, insurance navigation, sliding fee scale, same-day contraceptive services, and comprehensive approach to contraceptive care. Through trial and error, staff found digital media outreach that optimized Google search terms and radio advertising on selected stations to be somewhat effective in raising awareness and attracting new patients to the clinic.

Fourth, a policy environment unsupportive of women's reproductive health and political resistance to messaging about contraceptive choice constrained Washington University's outreach efforts. Given sensitivities about contraception and choice in the state's political environment, the awardee met with resistance to its messaging in some public venues (for example, the metropolitan transit system). The university also was reportedly reluctant to associate its corporate brand with suggestive messaging targeting young people. Acting on the advice of its community advisory board, C3 stopped using its full name and focused messages more on the clinic's accessibility and support services (expanded hours, same-day services, Spanish-speaking professionals, insurance assistance) than on specific contraceptive services. In interviews, staff also suggested that the political environment after the 2016 elections affected C3 enrollment in other ways. Women's fears of losing contraceptive coverage reportedly prompted a spike in interest in C3 services (and LARC methods, in particular). However, women from immigrant communities were reportedly less willing to seek services amid uncertainties about immigration enforcement policies.

> "The whole topic of women's health and birth control is still a bit taboo. Word of mouth is the best way to share information, but the topic of birth control and sex and sexual health is still a bit uncomfortable. That's a roadblock that we see."
>
> —C3 contraceptive counselor

Over time, C3 staff found word of mouth within patients' social networks to be the most effective means of recruiting new patients to the clinic, which staff attributed to patients' overwhelmingly positive experiences with C3 services. Similarly, staff found that C3 patients'

positive experiences made providers more willing to refer new patients in the future, which helped increase referrals over the three years of the HCIA R2 cooperative agreement.

2. Delivery of program services

a. Description of and changes to service delivery model

Washington University established the C3 clinic and delivered services in a manner consistent with the CHOICE service delivery model, which included patient education about contraceptive options, evidence-based contraceptive services (including same-day access to LARC), and follow-up support. To address patient barriers to care and adapt the model from a research setting to the real world, the awardee added insurance navigation and counseling as a core component of its C3 service delivery model.

The first step of the C3 service delivery process was designed to address financial barriers by giving patients accurate estimates of their likely out-of-pocket costs. This occurred before they received contraceptive counseling or clinical services so that patients could make informed decisions. Insurance coverage options for C3 services during the first two years of the cooperative agreement included Medicaid, the Women's Health Services Program (Missouri's Medicaid family planning waiver demonstration), and commercial insurance provided by an employer or purchased through the state's health insurance exchange. During the third year of the HCIA R2 cooperative agreement, Missouri suspended state participation in the federally funded Medicaid family planning waiver demonstration in favor of a state-funded program that does not cover or pay for services provided by an organization that also provides abortion services. Because Washington University provides abortion services, C3 services were not covered under the state-funded Uninsured Women's Health Services Program.

For patients not eligible for coverage under any of the insurance options, C3's status as a Title X clinic allowed it to provide family planning services at a reduced cost to patients (including undocumented immigrants) based on a sliding fee scale determined by patient income. In addition, women who applied for but were not yet enrolled in Medicaid could receive same-day services if they paid the sliding fee. Then, they would be retroactively reimbursed by Medicaid once they were enrolled. Commercially insured women could also choose to pay the sliding fee if it was cheaper than what they would be required to pay under their insurance plan.

The second step of the C3 service delivery process—a core feature of the C3 model—was to provide contraceptive counseling to all eligible patients at their first clinic visit, even if they already knew which method they planned to use. Non-clinician counselors were trained on and expected to follow standard protocols to deliver evidence-based content in a nondirective manner on the comparative effectiveness, risks, and benefits of contraceptive options to all patients before they made their decisions and consented to treatment.

As the third and final step of the service delivery model (once patients had received contraceptive counseling, decided on a contraceptive method, and consented to treatment), C3 clinicians were expected to conduct medical intake tasks and deliver the clinical services. Providing all contraceptive services during a single appointment was a central component of the C3 model that was intended to address a recognized barrier to women's adoption of more effective methods of contraception—namely, wait time. In other clinical settings, women who

wanted an intrauterine device (IUD) or implant typically had to wait for the device to be ordered and then return to the doctor's office or clinic for a second appointment to have it inserted. Many women often either failed to keep the second appointment or were encouraged to choose a less-effective, short-term method available at the time of their appointment. To avoid this, C3 sought to keep LARC devices in stock, taking advantage of Title X 340B pricing discounts to defray the high up-front costs that often deterred other practices from doing the same.

As a Title X clinic, C3 clinicians were also required to provide primary and preventive care related to women's reproductive health, in addition to contraceptive care. In response to recommendations from the state's Title X grantee organization, C3 also began providing preventive services related to men's reproductive health during the third program year.

The awardee hypothesized that adapting the CHOICE model to address the real-world constraints of third-party reimbursement would reduce barriers to evidence-based methods of contraception among women of childbearing age and among clinicians who provide family planning services. Washington University expected that this would increase the uptake of methods proven to be most effective and thus reduce unintended pregnancies and childbirth and their associated costs.

b. Evidence of service delivery effectiveness

Washington University was successful in its delivery of the C3 program. Specifically, the awardee (1) developed work processes that effectively supported delivery of the core intervention services of patient counseling, evidence-based education, and same-day service delivery; (2) implemented flexible staffing models and training to adapt to evolving service delivery needs; and (3) effectively engaged program participants in informed decision making for contraceptive care.

Delivery of intervention services. Washington University established the C3 clinic on schedule and delivered contraceptive care in a manner consistent with the service delivery model proposed in its HCIA R2 application. C3 staff implemented a workflow that supported the primary objective of reducing barriers to access and increasing uptake of effective contraceptive methods among women at risk for unintended pregnancy.

The awardee successfully addressed patients' financial barriers by providing insurance navigation, counseling, and support. The C3 social worker became a federally certified and state-licensed insurance navigator, who counseled uninsured patients about their options and helped them apply for coverage (depending upon their eligibility) through Medicaid, the Medicaid family planning waiver demonstration, or the state health insurance exchange. When Missouri withdrew from the federally funded Medicaid family planning waiver demonstration, the C3 insurance navigator continued to help uninsured patients apply for coverage under the state Uninsured Women's Health Services Program. However, staff reported that some women preferred to remain uninsured in order to ensure that they could be treated at C3 in the future. For patients who lacked insurance, C3 staff determined what their out-of-pocket payment would be under the Title X sliding fee schedule based on their income and household characteristics. By the third program year, the C3 social worker reported that insurance counseling and assistance had become a well-established routine: "I could do that in my sleep." In addition, C3 staff

reported working effectively as a team to gather the insurance and income data and give patients accurate estimates of their coverage and costs even before they came in for their appointment.

> "I think the patients really appreciate the time that we spend on the counseling part. Most of the places where they have been before, they don't explain anything to them."
>
> —C3 contraceptive counselor

In interviews, the C3 non-clinician contraceptive counselors reported that they reviewed contraceptive options with patients (in English or Spanish) and answered their questions before eliciting their preferences. The amount of time counselors spent and the level of detail they went into reportedly varied, depending upon patients' needs. However, counselors followed standard protocols in delivering evidence-based content to all patients. By design, counselors reported spending more time with adolescents and with women who appeared to have limited prior understanding of reproductive physiology or how different contraceptive methods worked. Anecdotally, counselors reported that patients often commented on the thoroughness of the C3 counseling, compared to that of other providers. The clinic's monthly self-monitoring statistics also indicated uniformly high levels of patient satisfaction with C3 counselors (10.5 points or higher in most months, out of 11 possible points).

C3 also implemented practices that expedited its ability to provide all contraceptive services during a single appointment. First, the awardee kept LARC devices in stock so that women who elected that option did not have to return to the clinic for a second appointment. Because many insurers (including Medicaid) required prior approval for LARC insertion, C3 staff also described making intensive efforts to secure prior approval before scheduled appointments in order to avoid delays in treatment. According to the awardee's monthly self-monitoring statistics, C3 succeeded in providing elected services during a single appointment 80 to 90 percent of the time across all program years.

Staffing and training. Washington University successfully recruited, hired, trained, and retained staff for the C3 program and implemented a flexible staffing model to adapt to evolving service delivery needs.

The awardee used a mix of new hires and existing departmental staff (including several who had worked previously with the CHOICE program) in a staffing model designed to match credentials to job functions efficiently. The model featured the use of non-clinician contraceptive counselors; a licensed social worker to provide insurance navigation and other social support; nonphysician advanced practice clinicians (nurse practitioners) to provide basic clinical services (including LARC insertion); board-certified OB/GYN physicians to provide supervision and backup clinical care; and administrative and other support staff to perform administrative, reporting, research, and medical assistance functions. With the exception of physicians and some research support staff who split their time between C3 and other clinical or research work, most staff serving in enrollment and service delivery roles worked full-time with C3. In interviews and in response to surveys, most staff reported having adequate support to do their jobs and manageable workloads. Staffing was stable, with very little turnover, until the closing months of the cooperative agreement.

As intended, staff training focused on credentialing the social worker to serve as a federally licensed and state-certified insurance navigator, evidence-based contraceptive counseling for

non-clinicians, informed consent and enrollment procedures, LARC insertion, electronic appointment and medical records systems, and data collection and reporting procedures and requirements. In response to survey questions, most non-clinical staff strongly agreed that training helped improve their job performance, although some thought that more training on insurance or billing requirements would have helped.

As enrollment increased, the awardee was able to continue the expanded evening and weekend hours and to prioritize fast-track appointment scheduling and patient care without adding new staff. To do so, the awardee made research and medical support staff available to assist with intake, scheduling, and precertification, when needed, and cross-trained staff to carry out those roles.

Recruitment and engagement of providers. Delivery of C3 services did not require the awardee to recruit or engage providers or provider organizations.

Engagement of program participants. Evidence suggests that the awardee was able to engage C3 patients in evidence-based decision making and adoption of the most effective contraceptive methods, which was critical to achieving its goal of reducing unintended pregnancies and their associated costs. According to C3 self-monitoring data, between 50 percent and 70 percent of C3 patients selected IUDs or implants each month—about twice the next-highest reported rate for a Title X clinic in Missouri and more than four times the average reported rate for all Title X patients in the state. Surveys conducted as part of routine follow-up with C3 enrollees also consistently showed high ratings of patient experience with the contraceptive counselors and clinicians. In interviews, C3 staff and leaders also cited the positive feedback they received from patients, the growing number of word-of-mouth referrals, and the unanticipated numbers of patients returning to the clinic for well-woman care as further evidence of C3's success in engaging its patients.

c. Barriers and facilitators associated with service delivery effectiveness

Several factors affected Washington University's ability to deliver intervention services, train and support C3 staff, and engage program participants. Above all, the awardee's prior experience with CHOICE gave it the basic model of service delivery that was essential to its effectiveness. However, the awardee encountered several barriers to service delivery, including (1) Missouri's decision not to expand Medicaid eligibility immediately prior to the program start and later decision to exclude the awardee from the state's Uninsured Women's Health Services Program; (2) Medicaid and other insurers' prior approval and other billing requirements; and (3) the data management and reporting requirements of C3 research, HCIA R2, Title X, and third-party payers. Below, we further describe these challenges and the factors that facilitated C3's ability to address them.

First, Missouri's limits on Medicaid eligibility and state policy on women's reproductive health services required the awardee to cobble together a viable support system to reduce women's financial barriers to access. In this context, C3's status as a Title X clinic was critical to its ability to provide services at reduced or no cost to patients who lacked insurance. Title X funding also gave C3 access to LARCs at a reduced price, which helped the clinic keep the devices in stock and available for same-day service delivery. However, Title X status also required C3 to provide a broader array of reproductive health services than those included in the

CHOICE model, including preventive services for men, which created new demands on service delivery. The awardee's increasing reliance on reimbursement from commercial payers also led C3 to expand its focus beyond contraceptive care, which added complexity to the task of differentiating its market niche.

> "We had no idea the complexity of all this insurance stuff. All we did with the CHOICE Project was birth control. Trying to merge all the women's health services programs and needs, colposcopies and Pap smears, along with still providing contraception has been a big learning curve for all of us."
>
> —C3 manager

Second, keeping up to speed with the complicated and changing insurance landscape in the unfamiliar territory of third-party billing requirements reportedly took its toll on C3 staff. In addition, Medicaid and other insurers' requirements for prior approval were potential barriers to providing same-day LARC services. To keep women from having to pay out of pocket for care or to delay care, C3 staff had to first confirm patients' insurance status and secure prior approval—steps that had not been necessary under the privately funded CHOICE model. C3 staff credited invaluable support from departmental billing staff and Washington University's billing infrastructure in helping them overcome this challenge, but acknowledged that it took time to master the nuances of the third-party prior approval and billing processes.

Third, collecting and managing the data needed to support C3 research, HCIA R2 and Title X reporting requirements, and third-party reimbursement meant that staff had to work across multiple systems and create some databases from scratch. Awardee leaders also acknowledged having underestimated the number of staff needed to carry out basic enrollment, registration, prior approval, documentation, and reporting tasks. In interviews, staff reported that C3 was able to address these challenges through cross-training and a flexible staffing approach. However, some strains in communication and teamwork were evident in the C3 self-monitoring data and in clinician and staff survey data collected as part of Mathematica's implementation evaluation.

Notwithstanding these challenges, the awardee's prior experience with the CHOICE program gave it the fundamental model of staffing, training, patient flow, and service delivery that was needed to effectively implement the C3 program. Staff reported few changes or challenges to the core work of contraceptive counseling and care. The level of commitment to the CHOICE model and C3's mission was also palpable in interviews with C3 staff at all levels.

C. Assessment of perceived program effects on the delivery of care and outcomes

Participating clinicians (nurse practitioners and physicians) and other C3 staff were consistently positive in their assessments of the program's effects on the delivery of patient care. All respondents to the clinician and staff surveys strongly agreed with the statement, "C3 is making a difference in meeting critical needs in our area/community." In interviews, staff emphasized that C3 provided needed services to women of limited means for whom avoiding unplanned pregnancy was a high priority. As evidence of its effectiveness in meeting this need, several staff cited Title X data showing that C3 patients were much more likely to adopt LARC or other highly effective contraceptive methods than patients in other Title X–funded family planning clinics in Missouri. "We're light years ahead," one interviewee said.

In surveys and interviews, almost all respondents cited buy-in from clinicians and staff and high levels of patient engagement as key factors associated with C3's perceived effects on the delivery of care. Virtually all staff and clinicians reported that members of the C3 program team collaborated effectively to provide high quality care and services and felt supported by management to do their jobs. In addition, nearly all staff and clinicians who completed the surveys indicated that they thought the program had positive impacts on care delivery and outcomes such as responding to patient needs in a timely way; providing services responsive to patient preferences, needs, and values; improving access to care for all patients; helping patients achieve their goals; and improving patients' quality of life.

In interviews, C3 leaders and staff said that they anticipated positive program outcomes similar to those achieved in the CHOICE project because C3 used the same program model of contraceptive counseling and care and experienced comparably high levels of LARC uptake. As part of the awardee's continuing research study, it will track C3 enrollees for at least one year and compare their outcomes to those of a similar group of Missouri Medicaid enrollees in order to assess the program's impacts on unintended pregnancy rates and costs. The awardee acknowledged that the absence of data on pregnancy intention for the comparison group was a limitation to its study design.

Although results of the C3 research study will not be available until a year after patient enrollment ends, the awardee conducted preliminary modeling of cost savings, comparing C3 participants' contraceptive use and unintended pregnancy rate to a usual care scenario that assumed patterns of contraceptive use and unintended pregnancy rates similar to those reported in national studies. The model showed a substantial reduction in the unintended pregnancy rate for C3 participants compared to the national rate among reproductive-age women. It also estimated substantial cost savings to Medicaid for maternity care and the first year of infant life, if the C3 model were extended to the Uninsured Women's Health Services Program in Missouri.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. Washington University's effectiveness in delivering C3 services—including, reducing barriers to access, engaging patients through evidence-based contraceptive counseling, and delivering clinical care with high rates of LARC uptake—make it likely that the C3 program reduced the rate of unintended pregnancy and its associated costs among C3 patients. However, several characteristics of the C3 program complicate the task of evaluating its impacts.

First, prior experience with the CHOICE program, although critical to the effectiveness of the awardee's implementation of C3, precludes identification of a baseline population lacking exposure to the intervention that could serve as a comparison group for a pre-post evaluation, because women in St. Louis had already been exposed to a similar intervention. Second, although the awardee has gathered data on pregnancy intention among C3 patients (because its objective is to decrease the rate of unintended pregnancy), no comparable data would be readily available for a comparison group of Medicaid patients because pregnancy intention is not captured in claims data. This would make it difficult to determine whether pregnancy outcomes in the comparison group were intended or unintended. Finally, although reduced pregnancy-

related costs may be discernible over the short term, the bigger potential impact of the C3 intervention relates to averted costs associated with childbirth, which would not be observable until at least one year after the delivery of services. For this reason, an evaluation of C3 impacts would need to track participants for at least one year after the delivery of program services.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of Washington University's 3C program.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Washington University received a three-month no-cost extension and continued to operate until November 30, 2017. The projected sample sizes in Table III.1 are based upon actual enrollment as of August 31, 2017.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: Washington University

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure	0[a]
Projected Medicaid population with 6 months of program exposure as of November 30, 2017	1,526[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	4,582
Likelihood of all-cause hospitalizations	2,932
MDE sample size requirement to detect 20% effect	
Total expenditures	1,146
Likelihood of all-cause hospitalizations	733
Participation/Selection bias of concern	Yes, patient self-selection high or high refusal rate
Intervention implemented that differs from baseline period	Questionable, patients may have been receiving intervention prior to HCIA R2 cooperative agreement
Claims sufficient to identify treatment and comparable comparison group?	No, low rate of identification of treatment group in baseline period using claims data with significant dilution of treatment effect within an intent-to-treat framework. Many will have no Medicaid baseline claims.
Likelihood of solid comparison group	Too early to determine due to delay in Medicaid data
Do claims identify the primary expected effects?	Some indirect effects may be observed in claims data, but important effects likely missing

Table III.1 *(continued)*

Evaluability domain	Response
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Lack of strong comparison group
Survey data for treatment group that will be analyzed	Staff, clinician, and beneficiary surveys
Implementation data that will be analyzed	We are working with the awardee to obtain its data on LARC use, unintended pregnancies, and unintended births.

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

At this point, we do not anticipate being able to conduct a rigorous impact analysis of the C3 program primarily because of significant challenges to constructing a strong comparison group. In particular, for any pre-period analysis, we cannot identify the beneficiaries in the treatment and potential comparison groups who were or were not in the Contraceptive CHOICE project—which continued to follow patients through 2013 and on which the current C3 program was modeled. We are concerned that the CHOICE project has likely contaminated the pool of participants and the baseline outcomes in Missouri. In addition, there are no comparable data sources on key outcomes used for the comparison group or in the pre-period treatment group, relative to the treatment group in the post-period. We are also concerned that the women who have Medicaid insurance may not have had it in the pre-period, given that the C3 program focuses partly on helping uninsured women enroll in Medicaid—which makes it difficult to track individuals over time in claims data and to isolate the impact of C3. Consequently, for this evaluation, we propose to carry out aggregate comparisons in key outcomes over time between program participants, women in Missouri and Illinois more broadly, and those in other geographic regions. This analysis will be done using nationally available survey data. Further, using awardee-provided data, we will report on LARC use, unintended pregnancies, and unintended births among treatment group beneficiaries. We will also report on the experiences of awardee staff, clinicians, and participants, based on our surveys.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

The awardee developed a model that would provide one bundled payment for a 90-day episode of contraceptive care, covering (1) initiation of the contraceptive method (including non-clinician contraceptive counseling, medical intake, clinician services, and pregnancy testing); (2) short-term follow-up and support; (3) an insertion fee modifier for LARCs (if applicable) (4) facility and administrative charges, (including insurance navigation and assistance); and (5) a dispensing fee for LARCs (if applicable). The awardee reported having had positive communications about the proposed payment model with Missouri HealthNet (the state Medicaid agency) and with a managed care payer. However, changes in state-level government, Medicaid management, and the Missouri Uninsured Women's Health Program slowed Washington University's progress.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

Washington University developed and submitted a proposal for a bundled payment model for a single episode of care (90 days) to Missouri HealthNet in December 2016. The model contains two reimbursement rates, one for LARC provision ($447.46) and one for shorter-acting contraceptive methods ($150.77). Both rates cover the initiation of the method (including contraceptive counseling), short-term follow-up to support continuation of the method, and facility and administration charges (including support from the social worker for patients who are trying to navigate their insurance options). The reimbursement rate for LARC also includes an insertion fee modifier for LARC and a dispensing fee for ordering, stocking, and inventory management of IUDs and implants at the health center. The actual cost of the contraceptive device is carved out of the bundle.

C. Status of the payment model

By working with Missouri's Medicaid program (Missouri HealthNet), Washington University has a potential opportunity to scale the model to other health centers. If Missouri HealthNet approved the model, the bundle would be available to all family planning providers who serve patients under traditional fee-for-service Medicaid. Missouri HealthNet would give Medicaid managed care plans until the next contract cycle to implement the new reimbursement model. In order to make the proposed payment model applicable to all providers serving Medicaid patients in Missouri, Washington University gained input from key stakeholders, reviewed internal costs such as staffing and facility fees, and reviewed Medicaid reimbursement schedules for contraceptive care in a number of states. Because this payment model is designed specifically for Missouri HealthNet, Washington University has begun to tailor it for other potential payers such as Medicaid managed care plans. In addition, the awardee has been conducting cost-effectiveness analyses by using claims data from Missouri HealthNet. In order

to assess the effectiveness of the program in reducing total cost of contraception and maternity-related services, the awardee planned to conduct a comparison of program enrollees and a matched cohort of nonparticipants.

Due to a change in administration in state government and significant turnover within Missouri HealthNet, the proposal submitted by Washington University is reportedly still under review. In the interim, Washington University had been talking to two Medicaid managed care plans. One Medicaid managed care plan had expressed enthusiasm about the innovative payment model and requested more information from the awardee. However, there were changes in the plan's leadership and thus Washington University had to identify new contacts and restart the conversation. The awardee planned to follow up and further communicate the value of the C3 program by submitting a detailed memorandum to introduce the bundled payment model and share results from its cost-effectiveness analyses. At one point, another Medicaid managed care plan was very interested in the program. However, its contract with Missouri HealthNet was not renewed. Awardee leaders mentioned that because Missouri Medicaid was going through an expansion to offer statewide managed care, there might have been competing priorities that contributed to the challenges in partnering with the Medicaid managed care plans. In addition, although stocking contraceptive devices is essential for providing timely services to patients, a dispensing fee is not currently covered by Missouri Medicaid managed care plans. In ongoing negotiations with Medicaid managed care plans, it was challenging for the awardee to gain support for a bundled payment model inclusive of a dispensing fee and it was unclear who would reimburse the dispensing fee.

D. Factors associated with the development of the payment model

In awardee documents and interviews, awardee leaders identified two primary facilitators of developing the payment model: (1) communication with payers and (2) consultations with the implementation and monitoring contractor. Awardee leaders also identified two key barriers to developing the payment model: (1) lack of timely feedback from payers and (2) a change in the status of the state's Uninsured Women's Health Services Program.

Awardee leaders said that developing close relationships with potential payers helped them better understand payers' needs and preferences for payment models. For example, the awardee developed the proposal for a bundled payment model based on feedback received from Missouri HealthNet. In addition, awardee leaders described input from the implementation and monitoring contractor as invaluable in developing the payment model and in preparing convincing payment memos for potential payers.

Awardee leaders also reported two key challenges to the payment model development. The first challenge, lack of timely feedback from payers, was evident when the awardee submitted a detailed memorandum for the proposed bundled payment model to Missouri HealthNet in December that was still under review nearly a year later. In part, this was due to changes in the leadership at Missouri HealthNet and a political climate in the state that made it challenging to get buy-in for contraceptive care. The second challenge was related to the status of the Uninsured Women's Health Services Program, in the wake of Missouri's departure from the federally funded Medicaid family planning waiver demonstration. The state-funded program continues to cover family planning services for uninsured women with incomes below 201 percent of the

federal poverty line, but the awardee doubts its ability to engage a program for which it is no longer an authorized provider. Although Washington University continued to see women covered under this program during the cooperative agreement, its services were not reimbursed.

This page has been left blank for double-sided copying.

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Washington University plans to sustain the C3 program after the HCIA R2 cooperative agreement ends. C3 leaders are assessing the staffing and budgetary requirements needed to continue the clinic's contraceptive counseling and clinical services, relying on funding streams from Title X and third-party billing for clinical services. Absent other sources of support, they foresee scaling back on evening and weekend hours, research support activities, and possibly on insurance navigation assistance. C3 leaders are also exploring ways to expand the program's scope of work, both as a means of enhancing revenues and in response to identified areas of need (for example, preventive services for the LGBTQ community). They are also interested in possibly extending C3 contraceptive services to underserved populations through the use of mobile units, but recognize that this would require new investments in research and development. If Missouri HealthNet pursues the bundled payment model for contraceptive care that the awardee has proposed, then core features of the C3 model could be replicated and reimbursed in other settings that serve Medicaid patients.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, Washington University began pursuing strategies to sustain the C3 program. The awardee was securing financial support for the program by (1) continuing membership in the Title X network to receive both 340B pricing for contraceptive devices and funding for clinical services; (2) maintaining a diverse payer mix, including commercially insured patients, to offset the costs of caring for uninsured and publicly insured patients; (3) serving an increasing number of return patients for reimbursable services beyond contraceptive care; and (4) developing a bundled payment model for approval from the state Medicaid program. The awardee also created a driver diagram to guide development of a detailed sustainability plan, estimated cost savings of C3 to the Medicaid program, and awaited Medicaid claims data to conduct a more rigorous comparative cost-effectiveness analysis.

The awardee had not scaled the program to other populations or areas, nor had other organizations replicated the program. However, if Missouri HealthNet approved the payment model, the awardee expected that it would apply to all Medicaid managed care plans across the state and to the state's Medicaid family planning waiver demonstration. All family planning providers who served Medicaid beneficiaries would be compensated for comprehensive contraceptive services and would therefore have the incentive to adopt core components of the C3 program.

C. Implementing the SSR plan: progress and changes

Sustainability. In the third program year, Washington University continued to focus on sustaining the C3 program and the staff remained committed to its mission. However, leaders also recognized that sustaining the program in its entirety was dependent upon the ability to obtain funding.

The awardee's long-term plan for sustaining the C3 model involved Missouri HealthNet's approval of its proposed bundled payment model and Medicaid managed care plans' adoption of a similar payment model. Washington University received a three-month, no-cost extension to its HCIA R2 cooperative agreement, to track research participants for longer (and thereby generate more robust impact estimates to share with insurers) and to pursue its payment model with interested payers. As discussed previously, the awardee reported that early discussions with the state's Medicaid leaders were promising, but their review and approval of the model was delayed after a change in the administration. The awardee also had earlier conversations about implementing the payment model with some Medicaid managed care organizations, but these discussions were also in hiatus because of changes in plan leadership and managed care contracts.

Over the short term, to keep the C3 program going in the absence of payment reform, Washington University was strategizing ways to generate revenues from other sources. The awardee expected to continue participating in the state's Title X network, which provided critical support for C3's clinical services. However, staff expressed concerns about the reliability of future Title X funding in the current policy environment. As a backup, they planned to obtain status as a Missouri sexually transmitted disease clinic, which would provide another route of access to 340B discounted pricing. The awardee also expected to maintain a payer mix that included commercially insured patients, whose higher levels of reimbursement helped offset

operating losses or lower levels of reimbursement from uninsured and publicly insured patients. However, the awardee was concerned that potential changes to the Affordable Care Act would limit contraceptive coverage for commercially insured patients. With the help of a strategic planning consultant, the awardee was exploring opportunities for expanding its billable services, as an additional means of continuing to support core C3 functions. Washington University also was exploring the possibility of fundraising with private donors. In August 2017, the awardee established an account with the Foundation for Barnes Jewish Hospital (the fundraising arm of Washington University's primary teaching hospital), which would allow C3 to accept donations with no additional overhead costs.

Overall, Washington University planned to maintain most of the core components of the C3 program. However, with the help of its strategic planning consultant, the awardee was considering alternative sustainability scenarios—based on projected numbers of new and returning patients, anticipated patient service revenue, and other sources of funding—and the extent of program and staff cuts that each scenario would entail. Based on these analyses, leaders thought that revenues generated from billing for clinical services would likely be adequate to support C3 clinical staff and services. However, they were less certain such revenues would support non-clinical staff or the administrative and operating overhead (notably, rent) previously covered as indirect costs under the cooperative agreement. In this context, awardee leaders were considering cross-training non-clinical staff to provide both contraceptive counseling and administrative and research support functions. They also noted that funding from other research grants (still pending at the time of our interviews) could help cover a portion of the salaries for research and administrative support staff, if some staff could split their time between C3 and other research.

C3 leaders planned to enroll patients through the end of the HCIA R2 cooperative agreement and hoped to track their outcomes for at least one year after enrollment, but they did not anticipate continuing C3's research component beyond that time without additional research funding. They also planned to retain the social worker over the coming year but were unsure about the continuing need for a dedicated insurance navigator, whose time in the past had mostly been spent helping patients enroll in the Medicaid family planning waiver demonstration, which was no longer a C3 option. Depending upon staffing levels, C3 leaders expected to reduce service hours (especially extended evening and Saturday hours). Marketing and outreach would also likely be curtailed after the end of the HCIA R2 cooperative agreement, absent additional funding. Some staff expressed concern that such reductions in capacity would reduce patient access as well as the number of patients that they could treat.

Scalability. In thinking about the future of C3, leaders were beginning to consider the possibility of developing mobile health units to extend services to some rural and underserved areas in Missouri, many of which reportedly lacked any OB/GYN providers. This would address a recognized need to reduce unplanned pregnancy across the state, but the leaders acknowledged that it would require a considerable investment of time and resources to develop, as well as new funding.

Meanwhile, C3 staff and leaders were considering other ways to expand the clinic's services, both as a means of generating revenues for the clinic and to address unmet needs of at-risk populations. Of particular interest were services to the LGBTQ community, although the

scope of services as they related to the Title X umbrella had yet to be determined. In the third program year, an intern trained C3 staff to provide care that was more "LGBTQ-friendly."

Although these plans represented a departure from C3's original mission, the awardee had already begun to move beyond its target population of women at risk for unintended pregnancy and beyond contraceptive care in order to support and sustain its core operations. For example, C3 already provided a broad spectrum of reproductive services and well-woman care and had begun to treat men. With the help of strategic planning and marketing consultants, the awardee was considering other ways of expanding C3's mission and services. This would likely entail changing the name of the clinic, although the awardee hoped to maintain the C3 acronym and brand.

Replicability. The C3 program had not been replicated at other sites. As noted in the Year 2 sustainability plan, if Missouri HealthNet approved the C3 payment model, the awardee expected that it would apply to all Medicaid managed care plans across the state. If adopted, this would create incentives for replicating core features of the program at other sites of care by covering the costs of contraceptive counseling by non-clinicians and stocking LARCs for same-day services.

D. Factors associated with progress toward implementing the SSR plan

Although the awardee considered adoption of the payment model vital to sustaining, scaling, and replicating the C3 model over the long term, strong patient volume and a diverse payer mix helped generate revenues to sustain the program over the short term. As the number of new patients increased during the award period, more patients than expected returned to C3 for routine care, which helped generate additional revenues. The percentage of commercially insured patients from the university community and elsewhere also was higher than expected, which helped to offset the costs of serving publicly insured patients and the uninsured. The awardee credited its marketing and strategic planning consultants with helping to increase C3's visibility and planned to use its online survey of community stakeholders to inform its sustainability planning. The awardee also expected the university's planned adoption of the EPIC electronic medical record system to facilitate patient scheduling and data collection for Title X reporting.

The state's decision to switch from the federally funded Women's Health Services Program to the state-funded Uninsured Women's Health Services Program was the biggest challenge to sustaining the program. Given Missouri's limits on Medicaid eligibility, the awardee had previously relied heavily on enrolling uninsured women in the federally funded Medicaid family planning waiver demonstration. Following the state's withdrawal from the federal demonstration during the third year of the HCIA R2 cooperative agreement, Washington University was no longer an authorized provider under the state-funded program enacted in its place and could not bill for services to this population. The awardee continued to provide most C3 services to patients previously covered under the program by using its HCIA R2 funding, but staff began to inform new and existing patients that they were not sure if they would be able to provide care to this group for much longer. The C3 insurance navigator offered to help uninsured patients apply for coverage under the state program, but staff reported that some women preferred to remain uninsured in order to be treated at C3 in the future. Staff acknowledged in interviews that it was not a sustainable strategy to provide services free of charge when the patients could receive coverage elsewhere. Although the Title X sliding scale could be used for uninsured patients and

commercial patients (if their situation warranted), staff believed that it could not be used if women had access to state-based insurance that would cover services elsewhere. This would require them to refer patients to authorized providers in the future.

Having sufficient staff capacity going forward was also a concern for the awardee, given the need to cross-train remaining staff to handle new roles. Program managers also noted that Washington University's generous vacation benefits made it challenging to cover clinic hours with a smaller staff, especially during summer months.

The awardee believed that the C3 model could be replicated in whole or part by other providers, if there were reimbursement incentives to do so. Even if the payment model were implemented, C3 leaders were not sure that insurers would include the proposed LARC dispensing fee. The awardee noted, however, that preliminary findings from another research study suggested that some clinics found it cost-effective to keep LARCs in stock even without a dispensing fee because it facilitated same-day contraceptive service. Without reimbursement incentives or a champion, clinics were unlikely to use the contraceptive counseling component of the C3 model, due to challenges identifying appropriate non-clinician staff to carry out the role, high rates of turnover among non-clinician support staff in most clinic settings, and the costs of training. Overall, the awardee believed that the reimbursement incentives incorporated in its proposed payment model would be critical to helping other clinics adopt the C3 service delivery model.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries as part of HCIA R2. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted in earlier reports, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA Policy Research

REPORT

HCIA Round Two Evaluation: Wisconsin Department of Health Services

April 18, 2018

Linda Barterian (Mathematica Policy Research)
Luke Horner (Mathematica Policy Research)
Kristin Lemos (Mathematica Policy Research)

Submitted to:
Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] The Wisconsin Department of Health Services received a 12-month extension ending August 31, 2018.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

B. Overview of the awardee's program

The Wisconsin Department of Health Services and its partners, the Children's Hospital of Wisconsin (CHW) and the University of Wisconsin Health–American Family Children's Hospital (AFCH), implemented the Special Needs Program (SNP)[4] to test a pediatric provider and reimbursement model for the care of children with medical complexity (CMC) and high resource utilization (Table I.1). CMC comprise the most medically complex subset of children with special health care needs whose families frequently face disjointed care and significant stress. The SNP sought to address the needs of CMC and their families by providing integrated health care including direct and consultative patient care, care management, and care coordination. The awardee's program goals included decreasing rates of hospitalization and emergency department (ED) visits, reducing the length of hospital stays, enhancing access to outpatient services, lowering health care costs, and improving satisfaction for families and primary care providers of CMC.

Using funding from HCIA R2, the awardee expanded and enhanced an existing SNP program at CHW first implemented as an inpatient service in 2002. AFCH launched a similar program in March 2014, but aligned with the SNP at the start of the cooperative agreement. The awardee reported that approximately 80 percent of children in these programs were enrolled in some form of Medicaid, which motivated the Wisconsin Department of Health Services to partner with the two hospitals. Launched on September 1, 2014, the enhanced SNP included two new elements: (1) expanded eligibility by broadening the criteria for medical complexity and including CMC with moderately high inpatient utilization[5] and (2) addition of staff members known as care coordination assistants (CCAs) to help coordinate care and provide emotional support to families. In addition, the awardee developed a payment model to ensure financial sustainability of the SNP at both institutions. Although funding for the SNP was originally scheduled to end August 21, 2017, the Wisconsin Department of Health Services received a 12-month no-cost extension from CMS.

[4] The Wisconsin Department of Health Services uses *SNP* to refer to the program at both CHW and AFCH. However, at AFCH, the program is called the Pediatric Complex Care Program.

[5] The original program implemented at CHW defined CMC as having conditions involving three or more organ systems that required the care of five or more specialists. Criteria for medical complexity for the SNP implemented under the cooperative agreement required the same number of affected organ systems but involvement from only three specialists. The original program targeted children with two or more hospitalizations totaling 10 or more hospital days, or with 20 or more clinic visits within a 12-month period. The enhanced SNP also included children with one or more hospitalizations of five or more hospital days, or 10 or more clinic visits within a 12-month period.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	The Wisconsin Department of Health Services (WI DHS) intends for the SNP to improve patient care, reduce avoidable health care utilization, and improve quality of life for participants and their families.
Major innovation	Building on an existing program at CHW, the SNP aimed to (1) provide direct and consultative patient care, care management, and care coordination to an expanded group of CMC including those with high ambulatory utilization and moderately high inpatient utilization; (2) implement the enhanced SNP at a second institution (AFCH); and (3) develop a payment model to sustain the SNP.
Program components	• Direct care provision • Care management • Care coordination (inpatient transitional and outpatient)
Target population	Children with medical complexity and high health care utilization in the year before enrollment
Theory of change/ theory of action	WI DHS and its partners hypothesize that enhanced care management and coordination for CMC will shift the burden of care coordination to the SNP, identify unmet needs, connect families to medical and social resources, and better coordinate care between primary care, specialty care, and inpatient care. This will lead to improved satisfaction among families and primary care providers and reduced health care utilization and spending.
Payment model	New Medicaid fee-for-service (FFS) payment for care management
Award amount	$9,444,864
Effective launch date[a]	September 1, 2014
Program setting	State Medicaid agency; two tertiary, acute care children's hospitals (CHW and AFCH)
Market area	CHW and AFCH are located in metropolitan areas (Milwaukee and Madison, WI, respectively). CHW serves patients from across WI, the Upper Peninsula of MI, and northern IL. AFCH serves patients throughout WI and from MN, IL, and IA.
Market location	Although both hospitals are located in cities, they treat patients from across the state and coordinate care with providers located near participants' homes.
Target outcomes	• Improve participants' quality of life • Improve primary care provider satisfaction • Improve family satisfaction • Reduce ED visits, hospitalizations, and total hospital days • Decrease total cost of care (9 percent reduction in per participant per month spending)

[a]After the initial planning period, the awardee's program became operational as of this date.

AFCH = American Family Children's Hospital; CHW = Children's Hospital of Wisconsin; ED = emergency department; SNP = Special Needs Program.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee successfully implemented its program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors: (1) enrollment, (2) service delivery, (3) staffing and training, (4) provider engagement, and (5) participant engagement. First, although the awardee initially underestimated the needs of participants in the expanded

target population and revised its enrollment target to approximately half of the original goal, it enrolled 1,062 participants—95 percent of its final enrollment target—by the end of the initial cooperative agreement. Second, the awardee provided direct care, care coordination, and care management services to participants throughout the cooperative agreement and improved timeliness and consistency of follow-up services when teams became fully staffed. Third, after encountering delays in hiring and training additional staff in Years 1 and 2 to meet higher-than-expected participant needs, the awardee reached staffing targets at the beginning of Year 3. Fourth, the awardee and its hospital partners reported strong engagement of providers on the program care team and referring providers who supported implementation of the SNP. Fifth, SNP program teams successfully engaged program participants, the SNP experienced low rates of withdrawal from the program throughout the three-year cooperative agreement, and SNP clinicians and staff perceived that the program had a positive effect on the delivery of care.

Impact evaluation. The impact evaluation in this narrative includes a description of the baseline characteristics of program participants. We are currently conducting a rigorous assessment of the impacts of Wisconsin Department of Health Service's SNP, and we will include this analysis in a future report.

Payment model. The Wisconsin Department of Health Services, a state Medicaid agency, developed a near-term payment approach for the SNP based on FFS reimbursement for targeted case management through a State Plan Amendment (SPA). Case management payments will cover program services not currently reimbursed under traditional Medicaid FFS and will include (1) a one-time payment for a comprehensive assessment and completion of a patient care plan at enrollment and (2) a monthly capitated payment for monitoring and follow-up activities. The Wisconsin Department of Health Services planned to submit the proposed SPA before the end of 2017 with rates effective retroactively to September 1, 2017. Because this model only included the participants enrolled in Medicaid, the awardee was also contemplating gaining participation from commercial payers.

Sustainability plans. The Wisconsin Department of Health Services and its hospital partners, CHW and AFCH, decided to use their current staff and other resources to sustain the SNP while they awaited approval of their proposed payment model. CHW had scaled its original program by broadening eligibility for the program under the cooperative agreement, and the payment model could promote scaling the program to additional children's hospitals with sufficient capabilities to implement the full model.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July to October 2016 with a sample of 25 potential respondents and achieved a response rate of 96 percent. The clinician survey was fielded from March to June 2017 with a sample of 18 potential respondents and achieved a 100 percent response rate. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a]Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

After reviewing the evidence available to us, we concluded that the awardee was successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

The SNP teams identified eligible children primarily through referrals from specialists, primary care providers, community programs, and children's caregivers. In addition, beginning in quarter 5 (September 2015–November 2015), program staff at CHW began to identify eligible children in hospital neonatal intensive care units (NICUs) and pediatric intensive care units (PICUs), meet their families, and invite them to participate in the program. AFCH staff also began

> "One of our nurse practitioners, she does NICU rounds once a week, and she's there at the rounds to help identify which of these kids that we feel will really be able to benefit from the program. Getting them in the program earlier rather than later, I think has been a big change."
>
> —Program leader

to advertise the program to inpatient units but did not recruit patients from the NICU because their hospital has a separate follow-up program serving this population.

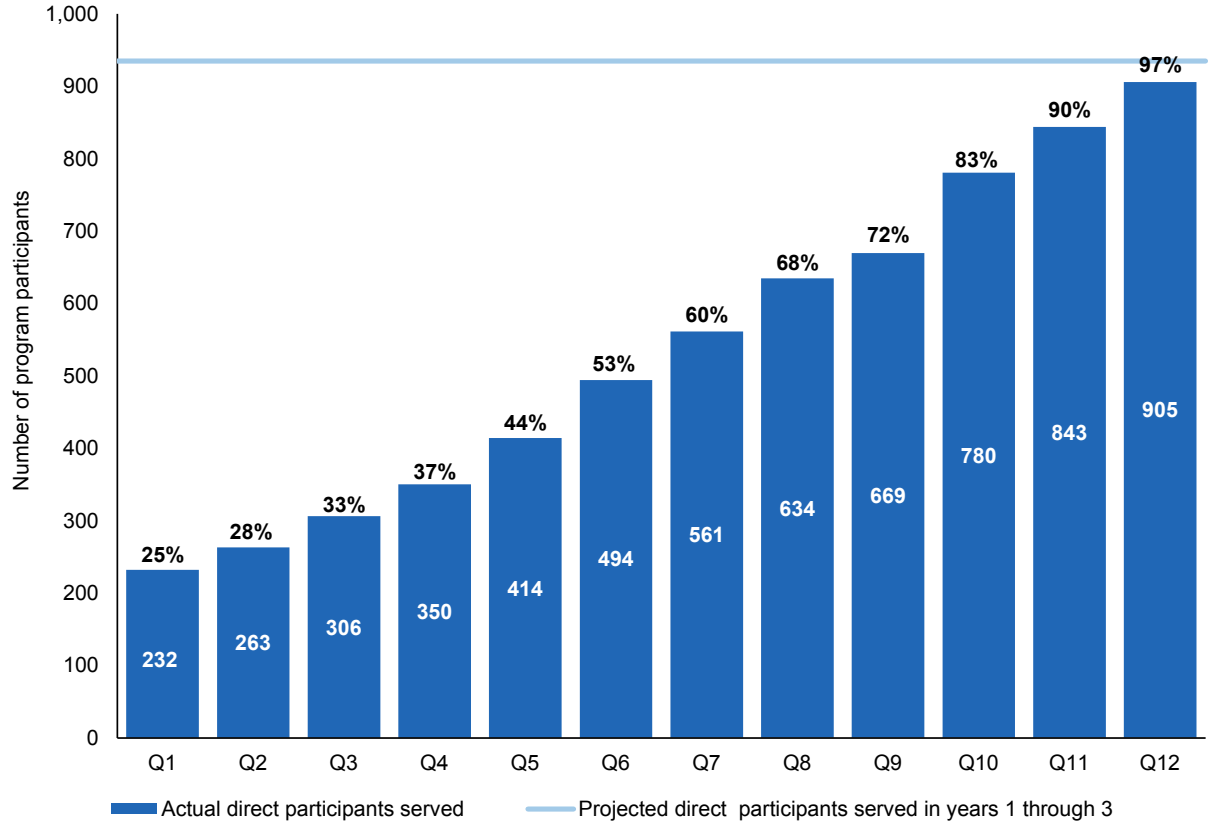
Program teams evaluated each referral to verify alignment with SNP eligibility criteria, and then a CCA called the family to schedule an assessment office visit before enrolling a child in the program. During assessment visits, the program team assessed patient needs, parental engagement and interest in the program, and potential patient eligibility for other Wisconsin Department of Health Services programs. After the assessment visit, program teams invited eligible children to enroll in the SNP, and parents decided whether to participate in the program. The program did not require formal consent documents. Program teams excluded a small number of children who they determined would not benefit from SNP—children with primarily mental health issues and children of families who do not want to participate in the program—as well as children and families whose needs they believed could be better met by a different program.

In Year 1, the Wisconsin Department of Health Services modified its original plan to enroll patients in two models of the SNP. The awardee originally proposed implementing an "intensive" model targeting children with high tertiary center use (two or more hospitalizations totaling 10 or more hospital days, or 20 or more outpatient clinic visits within a 12-month period) and an "ambulatory" model targeting children with moderately high tertiary center use (at least one hospitalization of 5 or more hospital days, or 10 or more outpatient clinic visits within a 12-month period). Program leaders anticipated that children enrolled in the ambulatory model would have less intense needs than children in the intensive model, so that the SNP could support more participants in that model with fewer staff. However, during the first year of the program, staff discovered that children in both models had similar levels of need and demand for services. In response, the awardee dropped the distinction in services for the two models during quarter 5 (September 2015–November 2015).

b. Evidence of enrollment effectiveness

The Wisconsin Department of Health Services reached 95 percent of its final cumulative projected enrollment of 1,121 by the end of the cooperative agreement. The awardee reported that it enrolled 905 children and youth on Medicaid (direct participants) and 157 on private insurance (indirect participants) from the launch of the program, September 2014, through August 2017 (Figures II.1 and II.2). However, the awardee revised its initial cumulative enrollment target of 2,040 participants (1,360 direct and 680 indirect) downward to 1,121 by the third program year. When measured against the three-year projection set at the end of Year 1 (enough time for the awardee to gain a better understanding of the number of eligible participants but before the awardee began planning for the end of their cooperative agreement), the awardee met 58 percent of its projection. The Wisconsin Department of Health Services decreased enrollment projections after discovering that children enrolled in the ambulatory model had similar needs as children in the intensive model. To address these higher-than-anticipated participant needs, the awardee adjusted staff-to-participant ratios and increased enrollment more gradually while hiring and training new staff. Notably, CHW reported 180 children were receiving care in the program at their hospital before the cooperative agreement and were enrolled in SNP during the cooperative agreement. For these children, the overall program was not new; only the modifications to the SNP funded by the cooperative agreement differed from their prior care.

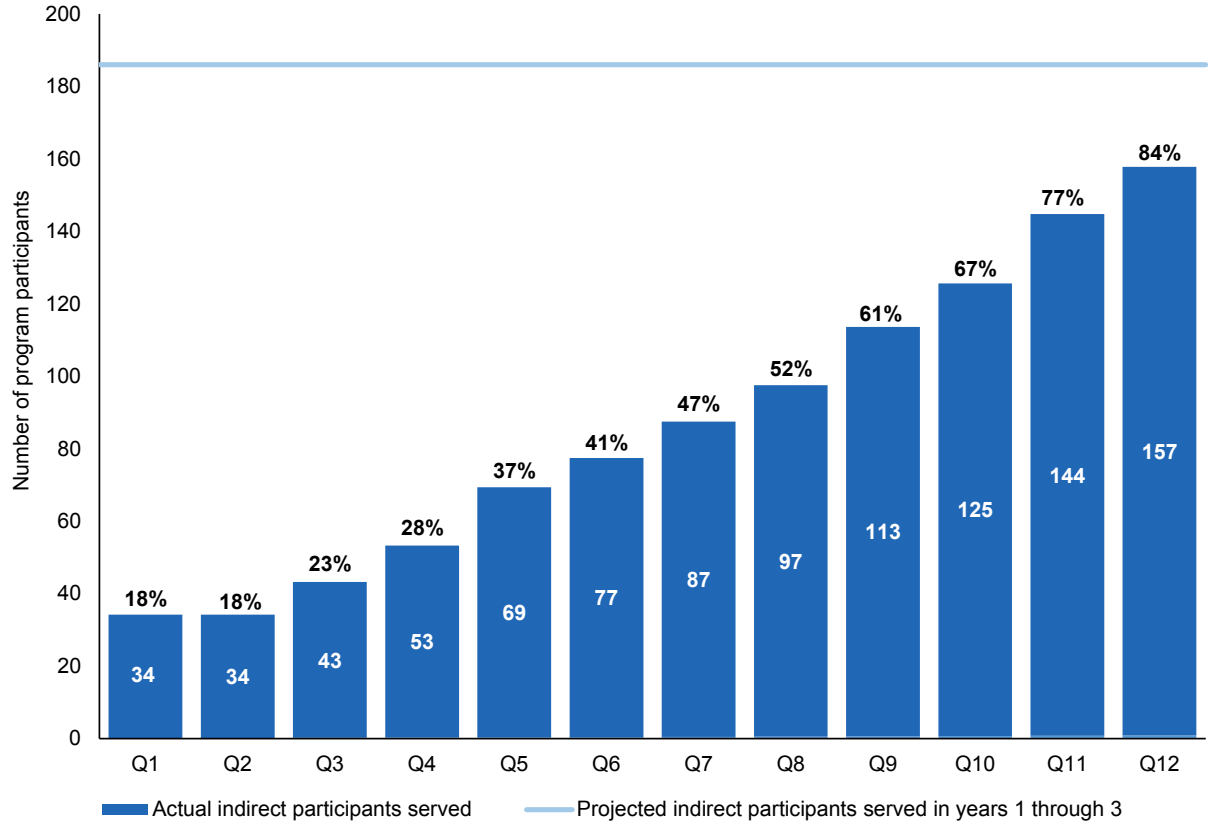
Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. Direct SNP participants were beneficiaries with Wisconsin Medicaid or the Children's Health Insurance Program as the primary or secondary payer. The awardee lowered its direct participant enrollment projection during the three-year cooperative agreement from 1,360 at the beginning of Year 1 to 1,317 at the beginning of Year 2 to 935 at the beginning of Year 3.

Figure II.2. Projected versus actual cumulative indirect participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected indirect participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Indirect program participants refers to the total number of unique participants who did not receive a service directly paid for by HCIA R2 funding, but who benefited indirectly from HCIA R2 funding support to providers. Indirect SNP participants were program participants who were not beneficiaries of Wisconsin Medicaid or the Children's Health Insurance Program.

c. Barriers and facilitators associated with enrollment effectiveness

Several factors influenced the Wisconsin Department of Health Services' progress in meeting its enrollment goals: (1) program teams' strong relationships with referral sources, (2) delays in reaching full staffing after adjusting staff-to-participant ratios, and (3) short-term staffing shortages resulting from occasional staff turnover and maternity leaves.

In interviews, SNP care team members at both hospitals reported that strong relationships with referring providers and the programs' positive reputation among providers and families led to a steady stream of referrals to the program over the course of the three-year agreement. CHW also worked to build relationships with NICU staff, which allowed program staff to supplement referrals with targeted outreach to eligible patients in the NICU. At AFCH, where the program was new at the beginning of the cooperative agreement, program leaders reported that referrals

increased. Staff attributed this increase to efforts to boost awareness of the program and its services among primary care and specialty providers and community organizations.

After the awardee adjusted nurse-to-participant ratios to accommodate higher-than-anticipated participant needs, leaders decreased enrollment goals to meet the revised staffing ratios. The teams encountered delays in hiring and training new staff, which limited their ability to enroll new participants at the pace originally planned. In interviews, awardee and program leaders reported that new nurses and CCAs needed time to learn the program and develop working relationships with other team members and departments before taking on a full patient caseload. Program staff also acknowledged that getting to know new participants and their needs required more staff time than anticipated. In addition, new participants had often experienced a medical event that triggered their enrollment and had high needs when they entered the program. Therefore, program teams limited the number of new participants assigned to new staff at one time and increased enrollment gradually as nurses and CCAs completed training. Notably, more than half of respondents of the clinician and staff surveys reported that unfilled positions and program staff turnover had presented barriers to achieving SNP goals.

Awardee leaders reported low staff turnover but acknowledged that occasional staff departures and maternity leaves placed added strain on program staff. Given the relatively small size of the program care teams, a staff departure or maternity leave stalled meeting enrollment targets. To help address this challenge, program leaders shifted responsibilities among team members and slowed enrollment of new patients temporarily.

Finally, the addition of CCAs to care teams under the cooperative agreement helped the awardee achieve SNP enrollment goals by increasing staff capacity to process referrals and schedule enrollment visits. When CCAs called patients to schedule enrollment visits, they explained the program and addressed patient transportation needs and other barriers, which helped reduce no-shows and increase enrollment.

2. Delivery of program services

a. Description of and changes to service delivery model

After a child enrolled in SNP, program teams created a customized care plan and assigned the child to a care team. SNP services included, at a minimum, (1) a follow-up phone call about two weeks after enrollment; (2) office visits at two months, four months, and every six months thereafter; (3) monthly care coordination telephone calls; (4) daily rounds whenever the child was in the hospital; and (5) post-discharge calls within 72 hours of a child's discharge from the hospital. Care teams had flexibility to adjust the mode and frequency of follow-up visits to meet participant needs. Participants remained enrolled as long as there was mutual agreement among program staff and families that SNP participation would continue to offer benefits.

Care teams consisted of physicians, nurse practitioners (NPs), registered nurse (RN) care coordinators, CCAs, and social workers. The teams also received support from administrative assistants. They provided SNP services as follows:

Direct and consultative patient care. Physicians and NPs provided direct patient care during scheduled clinic visits with participants, consulted in the inpatient setting, and supported

the development and implementation of patient care plans. They also provided care as needed when participants experienced medical events that could be managed with an urgent clinic visit.

Care management. SNP physicians and NPs worked with primary care providers and specialty providers to co-manage participants. This co-management consisted of collaborating on medical decision making with other physicians and participants' families and providing around-the-clock accessibility by phone to participants' families and other physicians, especially during acute changes in medical conditions and transitions of care between hospital units or from hospital to home.

Care coordination (outpatient and transitional). Care coordination teams consisting of an RN–CCA dyad ensured that participants' SNP and specialty follow-up appointments occurred in a timely manner, and served as the primary point of contact for patients' families. In addition, social workers addressed psychosocial, emotional, and socioeconomic issues that affected access and adherence to care—for example, by helping resolve insurance issues, educating families about government programs, and addressing transportation barriers to care. When a child was hospitalized, his or her care coordination team also conducted post-discharge follow-up phone calls.

During the fifth program quarter (September 2015–November 2015), the Wisconsin Department of Health Services made several changes to its original proposed service delivery approach. First, the awardee removed the distinction in services delivered to children enrolled in intensive and ambulatory models because SNP teams observed that children in both models had similar needs. Second, the awardee and its hospital partners had originally planned to classify participants into three tiers of care management and coordination based on variation in needs over time. Care teams had expected that after a child enrolled in SNP, the intensity of services required to meet the child's care needs would decrease with increasing duration in the program. However, they found that participants' health care needs fluctuated over time rather than tapered off in a predictable way. Therefore, the awardee dropped the use of tiers because they did not prove useful in predicting variation in participant needs for SNP services and informing staffing and caseload estimates.

b. Evidence of service delivery effectiveness

Overall, the Wisconsin Department of Health Services' hospital partners were successful in delivering program services. Program teams at CHW and AFCH reported providing program services including direct care, care management, and care coordination throughout the cooperative agreement, although they did not consistently implement follow-up calls within 72 hours after discharge until care teams became fully staffed. The awardee originally underestimated the needs of SNP participants and encountered delays in hiring and training staff but later achieved staffing targets at the beginning of Year 3. Leaders and staff at both sites reported a high level of engagement of providers from the program care team and strong relationships with referring providers who had positive perceptions of the program and its benefits for CMC and their families. Finally, both program sites had low rates of withdrawal throughout the three-year cooperative agreement, suggesting the teams successfully engaged families after their child enrolled in SNP.

Delivery of intervention services. SNP care teams delivered program services largely as intended while gradually increasing enrollment. They reported completing follow-up visits and monthly phone contacts consistently for enrolled participants. Program staff reported they adhered to the SNP service delivery model at a minimum, sometimes adapting timing or frequency of follow-up to meet increased patient needs. For example, in some cases, staff adjusted timing of follow-up visits in the clinic to coordinate with a patient's appointments with specialists, minimizing the travel burden on patients and families. However, care team staff also reported that before reaching full staffing in Year 3, they did not consistently complete follow-up calls to SNP participants within 72 hours of their discharge from the hospital. According to the awardee's self-reported monitoring data, the rate of post-discharge follow-up phone calls completed within 72 hours of discharge remained at 52 percent or below at CHW and 69 percent or below at AFCH during Year 2. After hiring and training new CCAs and achieving staffing targets at the beginning of Year 3, program teams improved timeliness and consistency of follow-up phone calls and appointment scheduling. The awardee's self-reported monitoring data for Year 3 showed an 85 percent and 89 percent average quarterly rate of post-discharge follow-up phone calls (completed or attempted) at CHW and AFCH, respectively.

Staffing and training. After encountering delays in hiring staff to accommodate an adjustment in the SNP staff-to-participant ratio from 1 RN care coordinator to 500 children to 1 RN care coordinator to 250 children in Year 1, the Wisconsin Department of Health Services and its hospital partners achieved staffing goals for the SNP at the beginning of Year 3. From the project's inception through the end of Year 3, the awardee hired 37.81 full-time equivalent staff, which is equivalent to 107 percent of its three-year projection. This included CCAs, RNs, NPs, physicians, administrative staff, social workers, and other support staff (Table II.2). The awardee also retained a high level of SNP staff throughout the three-year cooperative agreement, retaining an average of 97 percent of staff per quarter and never falling below 93 percent in a quarter.

Table II.2. Wisconsin Department of Health Services staffing totals and new hires from project inception through quarter 12

Program staff role	Total staff (number of individuals)	Total staff (full time equivalents)
Care coordinator/case manager/patient navigator	12	10.96
Management or administrative staff	6	2.82
RNs	12	10.63
NPs	12	7.97
Physicians	5	3.24
Social workers	3	1.26
Other	2 (research coordinator, medical assistant)	0.93 (research coordinator, staff trainer, medical assistant)

Source: Implementation and monitoring contractor, Quarterly Awardee Performance Report: Wisconsin Department of Health Services, 12th quarterly reporting period (June–August 2017).

NP = nurse practitioner; RN = registered nurse.

The Wisconsin Department of Health Services reported training 45 staff members from the program's inception, meeting its training projection for the three-year cooperative agreement. Training for physicians, nurses, NPs, and CCAs included orientation and shadowing SNP staff.

Social workers, administrative staff, and the medical assistant also shadowed SNP staff, met with staff from related departments, and completed internal hospital competencies. Consistent with these training efforts reported by care teams in interviews, respondents of the non-clinician staff survey reported participating in a broad range of informal training activities including shadowing other staff (96 percent), mentoring (82 percent), individual supervision (83 percent), and huddles (88 percent). Nearly all survey respondents reported that formal training helped them improve their job performance within the program (56 percent strongly agreed and 39 percent somewhat agreed) and that they learned new skills important for their role (50 percent strongly agreed and 44 percent somewhat agreed).

Engagement of providers. The awardee and its hospital partners reported strong engagement of providers from the program care team and referring providers that supported implementation of the SNP. CHW and AFCH care team providers had delivered services to CMC as part of their respective programs before the cooperative agreement, and program leaders and staff reported sustained high levels of provider commitment to the SNP. All respondents of the clinician survey reported they were attracted to SNP because of its purpose or mission and because they thought the program would improve access to care or improve patient care. Nearly all of these respondents also thought the effort to implement the SNP has been worth it to them (94 percent) and indicated satisfaction with their current role in the program (50 percent very satisfied, 44 percent somewhat satisfied). Leaders and staff at both sites also reported strong relationships with referring providers who had positive perceptions of the program and its benefits for CMC and their families. CHW had established strong referral relationships before the cooperative agreement that led to steady referrals throughout the agreement. AFCH staff described their efforts to build a close collaboration with their hospital's palliative care team to reduce duplication of services and ensure role clarity between the two groups.

Engagement of program participants. SNP teams reported successfully engaging participants throughout the three-year cooperative agreement. In interviews, program leaders, staff, and clinicians reported high levels of participant engagement in the SNP. In addition, 96 percent of respondents of the staff survey and 95 percent of respondents of the clinician survey agreed the program successfully engaged participants. During interviews, care teams noted that monthly follow-up telephone calls helped maintain engagement with patients and families between clinic visits. Both sites had low rates of participant withdrawal, and program leaders reported that patients and families who initially declined to join the program often later chose to enroll. The awardee's self-monitoring data show rates of participant withdrawal and loss-to-follow up of one percent or below at both hospitals throughout the cooperative agreement. Finally, 78 percent of respondents of the clinician survey indicated that participant engagement was the most helpful factor in the SNP achieving its goals.

> "We tell families that we are there for their support system and to help advocate for their child in the health system. That is a huge asset that we have with our families. You can see at first maybe they're hesitant. But then, once that first time you help them or they see the positivity of the program, then they're contacting you a lot more….There are always going to be some people that are harder to engage. I think it happens pretty infrequently."
>
> —Program staff member

c. Barriers and facilitators associated with service delivery effectiveness

Two key factors presented challenges for the Wisconsin Department of Health Services and its hospital partners in implementing the SNP: (1) finding staff with the appropriate mix of skills and fit for the program and (2) hiring and training new staff while simultaneously increasing program enrollment.

SNP teams struggled to find qualified candidates at all positions on care teams. In interviews, program leaders and staff attributed recruitment challenges to the unique skill set required for SNP positions. After experiencing some staff turnover in Year 1, program teams learned that candidates with the best fit for the SNP had strong communication skills, familiarity with the hospital system, a commitment to teamwork, and a passion for the CMC population. Because of the diversity of conditions among CMC, the SNP requires medical staff with broad expertise and the ability to work closely with other providers to meet participants' needs. Program teams adopted team-based approaches to evaluate candidates, and CHW used a parent panel to interview candidate physicians to ensure they understood the family-centered culture of the SNP. Added challenges in recruiting physicians and NPs included a requirement to work one weekend each month and less competitive salaries compared to other specialties. However, staff also reported that the positive reputation of the program and collaborative team environment helped attract interested candidates for all positions. In particular, both programs reported receiving many strong candidates for the CCA position, which provided a good opportunity for advancement, as one program leader described: "[The CCA position is] a good 'step up' from a certified nursing assistant or other clinic-based administrative aides so it draws applicants looking to move up in their field."

SNP teams faced challenges in bringing new staff up to speed while addressing the needs of an influx of new patients who required more time and attention than existing patients. Training new staff to care for the SNP population with highly variable needs took as long as six months. Both programs gradually increased caseloads for new staff and ensured a balance between patients new to the program and patients who have participated for six months or longer to help them reach target caseloads and provide better care.

> "It can be challenging. You're throwing a nurse and a CCA that know nothing about each other to essentially work 40 hours a week together. When new CCAs come on, we try really hard to make sure that the nurse and CCA early on are in really close communication. We'll tell the nurse, 'If you're going to see patients, that's priority over me teaching them how to schedule. We'll get to that another day.' It's vital that the CCA meets those patients right away so that they can start to put a face to the name. It builds a bond with the CCA and nurse as well early on."
>
> —Program CCA

Several factors facilitated implementation of the SNP: (1) teamwork and collaboration among care team members, (2) CCA support for administrative and care coordination tasks, and (3) support for the program among awardees and hospital partner leaders.

Program leaders and staff at both sites highlighted the importance of a team-oriented culture and shared commitment to learning and quality improvement in supporting SNP implementation. One staff member explained, "We definitely support each other. I think that any of us that see someone's really struggling, we're willing to help each other out." Within the broader awardee team, leaders from AFCH and CHW cited being "on the same page" from the beginning, which

helped them share approaches with each other and maintain alignment between the programs. For example, AFCH visited CHW to see how its CCAs interacted with other program staff in preparation for adding the role locally. Similarly, respondents of the clinician survey reported that members of the program team collaborate to effectively meet participant needs (66 percent strongly agreed and 33 somewhat agreed). Among respondents of the staff survey, 92 percent strongly agreed that members of the program team collaborate effectively to meet participant needs, 79 percent were very satisfied in their current role for SNP, and 100 percent felt supported by colleagues and management to do their job for SNP (75 percent strongly agreed and 25 percent somewhat agreed).

The addition of the CCA role to the SNP care team of physicians, NPs, and nurse care coordinators from the beginning of the program enabled care teams to deliver more timely and proactive care. Program teams assigned responsibility for administrative tasks to CCAs, which allowed nurses to expand their caseloads and increase their accessibility to patients. In addition, CCAs conducted many of the monthly follow-up calls with participants' families. Notably, at CHW where the SNP program originated, members of the care team described some mild "growing pains" in adjusting to changes in the care team to expand the program under the cooperative agreement. Although care teams highly valued the support of CCAs in delivering high quality care and serving more children, they needed to make adjustments to expand the program from approximately 200 children to nearly 600 children. Program leaders and staff acknowledged that efforts to increase efficiency required changing care team roles and resulted in participants being less familiar with individual staff members.

> "Since we've had the CCAs on board, we can do a lot more administrative calls to families for appointment reminders. Part of being in the program is at least a once monthly contact by the program, even if your child is well....That's always been a goal of the program, but it was very difficult for the nurses to keep up with that. I think we've gone from about 80 percent contact with every family to almost 100 percent."
>
> —Program care team member

Program leaders at the Wisconsin Department of Health Services and both hospitals reported strong institutional and leadership support for the SNP. SNP leaders at the Wisconsin Department of Health Services expressed a strong commitment to addressing the needs of CMC and reducing fragmented care that could result in adverse outcomes for these children. Hospital program leaders also described institutional support for the SNP, including financial support to cover the costs of participation among non-Wisconsin Medicaid participants. In interviews, the awardee and leaders at both hospitals reported having prominent program champions.

C. Assessment of perceived program effects on the delivery of care and outcomes

Seventy-eight percent and 65 percent of surveyed clinicians and staff, respectively, indicated that the SNP has been very effective in achieving its goals. In interviews, the Wisconsin Department of Health Services and SNP program leaders and staff at both sites reported that SNP goals—to reduce and shorten hospitalizations, reduce ED visits, enhance access to outpatient services, and lower the cost of care for CMC through improved quality of care—remain attainable.

Similarly, SNP clinicians and staff expressed positive perceptions of the program's effects on care delivery and participants' satisfaction and quality of life (Table II.3). In addition, 83 percent of clinicians and 96 percent of staff responding to the survey strongly agreed that the SNP was making a difference in meeting critical needs in the area/community. Also, program staff and leaders shared in interviews that they have observed improvements in quality of life for participants' families and heard parents express relief when the SNP team coordinated care for their child, which often took an immense amount of parents' time and energy.

> "We do have a couple of parents who had to leave the work force because of their child's needs who have returned to the workforce because we've helped to stabilize their child... Some families who were on the verge of either needing to quit or get fired from their job because of missing so much work due to their child health needs, enabling them to stay in the work force is helpful. Certainly, we work with the schools closely to talk to management so that [participating children] miss the least amount of school possible. It certainly feels anecdotally that we are having an impact there."
>
> —Program leader

Table II.3. SNP clinician and staff perceptions of the program's effects on participants and delivery of care

Do you believe SNP has had a positive impact, negative impact, or no impact on:	Positive impact (percentage)	
	Clinicians	Staff
The quality of care and services you provide to participants	100	96
Your ability to respond in a timely way to participant needs	100	83
Your ability to provide care or services that are responsive to participant preferences, needs, and values	100	92
Access to care or services for all participants	94	92
Achievement of participants' health goals	94	83
Participant satisfaction	94	100
Participant quality of life	89	96
Care coordination	100	96

Source: HCIA R2 evaluation surveys of awardee's non-clinician staff (24 respondents) from July to October 2016 and clinicians (18 respondents) from March to June 2017.

SNP = Special Needs Program.

SNP leaders and staff reported in interviews a strong sense of teamwork and mutual support among program team members. In addition, all respondents of the clinician survey indicated they had observed better team interactions, an improved work environment, and improved participant care as a result of participating in SNP. Among respondents of the staff survey, 79 percent indicated they were very satisfied in their current role in the SNP, and 83 percent strongly agreed they were an important part of the program team. However, among survey respondents, 56 percent of clinicians and 33 percent of staff indicated administrative requirements including documentation for the SNP took up too much time. In interviews, program staff reported that documentation sometimes suffered when they strived to meet patient needs with limited staffing capacity.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis. The Wisconsin Department of Health Services and its partners implemented the SNP to improve the quality of care for CMC and improve outcomes for participating children and their families. Several SNP implementation issues have implications for interpreting and analyzing program impacts, including the following:

SNP did not near its enrollment target until the beginning of Year 3, so exposure to the intervention was 12 or fewer months for many participants. Program leaders estimated a minimum of 18 months of participation in the SNP to observe expected outcomes. Because the program design changed in Year 1 and the program did not reach full staffing until the beginning of Year 3, the SNP did not near its enrollment target of 1,121 until Year 3. Because 31 percent of the cumulative total enrollment occurred in Year 3, many SNP participants had participated in the program for 12 or fewer months by the end of Year 3. Therefore, exposure to the intervention for many participants might not be long enough to detect potential changes in utilization and costs of care.

For children who were in the SNP at CHW before the cooperative agreement, changes in utilization and costs reflect longer exposure to the basic intervention but likely represent only the marginal effects of the changes to the program under HCIA R2. At CHW, 180 children, representing 17 percent of cumulative total enrollment, had participated in the hospital's program for CMC before participating in SNP under the cooperative agreement. With low graduation and withdrawal rates, many of these children might have received services for more than 36 months. Rollover of CMC from a prior program into SNP was less of an issue at AFCH because its hospital's program had started only a few months before the cooperative agreement.

Acute care utilization and total costs might increase in the short term as SNP teams work to meet intense or previously unmet needs of new participants. Program leaders and staff reported that children often enter the SNP with intense medical needs and require six or more months in the program before their conditions stabilize. One program leader explained, "During that first year, there actually may be increased spending because they've been disconnected from all these different services or we recognized that the parents are on the verge of a breakdown and get home nursing involved in the care. Additionally, we have many patients that are referred to us in advance of a large surgery because surgeons have seen the value of the work we do and want us to be involved in managing their patients."

Some CMC have illnesses that get progressively worse, which could limit the potential for reducing acute care utilization. In these cases, the SNP aims to reduce care burden on families and improve participants' quality of life and coordination of care.

The SNP aims to improve satisfaction and quality of life for participants and their families, and these impacts are more difficult to quantify. We cannot observe these important potential impacts in the data available to us that includes a potential comparison group, specifically Medicaid administrative and claims data.

This page has been left blank for double-sided copying.

III. FINDINGS FROM THE IMPACT EVALUATION

This chapter presents an update to our assessment of the evaluability of the Wisconsin Department of Health Services' program and baseline characteristics of the treatment group.

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: Wisconsin Department of Health Services

Evaluability domain	Response
Projected number of Medicare fee-for-service (FFS) enrollees with 6 months of program exposure by February 28, 2018	Not applicable
Projected number of Medicaid enrollees with 6 months of program exposure by February 28, 2018	908[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,649
Likelihood of all-cause hospitalizations	380
MDE sample size requirement to detect 20% effect	
Total expenditures	412
Likelihood of all-cause hospitalizations	95
Participation/Selection bias of concern	Yes, provider clinical judgment/non-claims data used to identify treatment group
Intervention implemented that differs from baseline period	Questionable, patients may have been receiving intervention prior to HCIA R2 cooperative agreement
Claims sufficient to identify treatment and comparable comparison group?	Yes, high rate of identification of treatment group with claims data within an intent-to-treat framework
Likelihood of solid comparison group	Some issues, but probably surmountable; expect to select a comparison group
Do claims identify the primary expected effects?	Some effects observed in claims data, but important effects likely missing
Core outcomes estimation method	Difference in differences
Primary reason for no rigorous evaluation	Not applicable
Survey data for treatment group that will be analyzed	Clinician, staff, and beneficiary surveys
Implementation data that will be analyzed	Time spent on care coordination activities

[a]The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

We anticipate constructing a valid comparison group composed of beneficiaries who are similar to beneficiaries in the treatment group (that is, who are also children with medically complex conditions) but who live far from the two participating hospitals. We can come close to replicating the awardee's eligibility criteria by using claims-based algorithms, though we may have some selection bias issues because providers may have input into which beneficiaries are enrolled. Our projected final analysis sample (shown in Table III.1) is based on the number of treatment group beneficiaries when enrollment ended in August 2017; the analysis sample will likely be large enough to detect plausible effects for the full sample, although we may not have sufficient sample size to detect effects separately for the two sites.

B. Characteristics of Medicaid participants at baseline

This section presents the first summary of the treatment group's baseline sociodemographic characteristics and medical conditions that was previously included in the second annual report. We measured both by using Medicaid eligibility and claims data for the year before the date on which each beneficiary enrolled in the enhanced SNP. Eligible participants enroll in the enhanced SNP after being referred by physicians, primary care providers, or community programs, or when they are recruited by program staff from PICUs or NICUs. The treatment group assessed for this analysis includes Medicaid beneficiaries concurrently enrolled in the enhanced SNP according to lists from the awardee. We plan to analyze Medicaid expenditures and service utilization after we receive Medicaid data that would cover the baseline period and a minimum of a six-month follow-up period for all of the program enrollees.

1. Baseline characteristics of the treatment group

The awardee began to enroll children in the enhanced SNP on September 1, 2014.[6] As of the end of August 2017, 908 Medicaid beneficiaries had been enrolled in the program since the enhanced SNP launch date.[7] This group makes up an estimated 85 percent of program participants. The awardee defines these participants as direct participants. The remaining 15 percent of program participants enrolled since the enhanced SNP launch date (an estimated 160

[6] The SNP has operated continuously since 2002; with HCIA R2 funding, the awardee has enhanced the existing SNP at CHW and expanded the enhanced SNP to a second site at the AFCH in Madison. The enhanced SNP model, launched on September 1, 2014, expanded eligibility by broadening the criteria for medical complexity and including CMC with moderately high inpatient utilization and added care coordination assistants (CCAs) to help coordinate care and provide emotional support to families. Participants who first enrolled in the SNP between its initial launch in 2002 and August 31, 2014, were excluded from this analysis.

[7] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

children[8]) received enhanced SNP services not funded by the cooperative agreement; the awardee defines these participants as indirect participants, and they are not included in the evaluation.[9] These participants are predominantly privately insured. Participants who first enrolled in the program before the launch of the enhanced SNP under the cooperative agreement were excluded from this analysis.

In presenting the baseline characteristics, we restrict the treatment group to program participants who were enrolled in Medicaid when their eligibility for awardee-provided services began (that is, their program enrollment date) and who were enrolled in Medicaid for a total of 90 days during the baseline year (the 365 days immediately before the date on which they enrolled). The calendar period covered by the baseline quarters is determined by the enrollment date for each participant and therefore varies by participant. Furthermore, we restricted the analysis to participants enrolled on or after the cooperative agreement's enrollment date (September 1, 2014) through December 31, 2015 (the end of the period for which Medicaid data were available at the time of this analysis). After we excluded participants who did not meet the above criteria (which includes 15 participants enrolled between September 1, 2014, and December 31, 2015, who did not fully meet the Medicaid eligibility criteria and an additional 690 participants first enrolled in the program after December 31, 2015), a total of 203 Medicaid-eligible program participants (referred to hereafter as "beneficiaries") are included in the analysis of baseline characteristics for this report.

2. Baseline characteristics for the full set of Medicaid beneficiaries

Our analysis of baseline demographic characteristics indicates that the awardee is recruiting a population consisting primarily of infants and very young children, most of whom are receiving Supplemental Security Income (SSI) benefits and who have complex medical conditions. Half (50 percent) of beneficiaries are age 2 or younger, and an additional 16 percent are age 3 to 5; conversely, just 18 percent are ages 11 to 17 (Table III.2). Beneficiaries are slightly more likely to be male (54 percent) than female. Unfortunately, the race and ethnicity of nearly two-thirds (62 percent) of beneficiaries is unknown. Among beneficiaries for whom data are available, 63 percent are white, whereas 18 percent are black. It is likely that this reflects the racial distribution of Wisconsin's Medicaid population, where 64 percent are white and 17 percent are black (based on data for whom race data are available in the MAX 2012 files). One-third (32 percent) of beneficiaries reside in the Milwaukee metropolitan area (where CHW is located), and 39 percent reside in other metropolitan areas in the state, including Madison (where the AFCH is located). Just 28 percent reside in rural areas. Seventy-eight percent of beneficiaries receive SSI benefits, which is substantially more than the statewide rate of SSI receipt among child Medicaid beneficiaries of about 5 percent (based on MAX 2012 data). Approximately 7

[8] In its quarterly reports to the implementation and monitoring contractor, the awardee reports total program enrollment, which includes participants first enrolled before the enhanced SNP launch date; these participants were excluded from the evaluation. Therefore, the number of indirect participants enrolled since the enhanced SNP launch date is an estimate based on information from the awardee regarding the percentage of children enrolled in the program who are direct versus indirect participants.

[9] Indirect participants are not included in the awardee's quarterly finder file.

percent of beneficiaries are in foster care (which is not necessarily mutually exclusive of SSI benefit receipt). A relatively small number (9 percent) of beneficiaries are enrolled in a managed care plan, and the remainder receive care on a FFS basis, likely reflecting the significant health care needs of this population. Finally, for 41 percent of beneficiaries, a third-party insurance plan is their primary payer, as defined by the presence of any third-party liability (TPL) payment in the Medicaid claims data during the beneficiary's baseline year.

The analysis of claims-based characteristics suggests that the awardee is successfully recruiting beneficiaries who have complex chronic conditions (93 percent of beneficiaries) (Table III.2).[10] This is expected, given that the awardee's eligibility criteria for medical complexity are somewhat more restrictive than that used by the pediatric medical complexity algorithm (PMCA).[11] The most common diagnoses among beneficiaries include neurological diagnoses (present in 63 percent of beneficiaries), gastrointestinal diagnoses (present in 49 percent), pulmonary-respiratory diagnoses (present in 46 percent), cardiac diagnoses (present in 45 percent), and mental health diagnoses (present in 45 percent).[12] We plan to use the PMCA's condition category flags as a starting point to identify beneficiaries with medical complexity for the comparison group. Along with the flags, we will incorporate other criteria, including a claims-based count of the number of specialists beneficiaries are seeing. We expect to identify beneficiaries with medical fragility by using the utilization-based selection criteria defined by the awardee.

[10] We identified the conditions by using both diagnoses available in Medicaid claims data for up to three years before program enrollment (if available) and the pediatric medical complexity algorithm.

[11] The PMCA defines complex chronic health conditions as those meeting at least one of the following three criteria: (1) more than one body system is involved, and each must be indicated in more than one claim; (2) one or more conditions are progressive; and (3) one or more conditions are malignant. For comparison, the awardee defines CMC as children who have chronic conditions involving three or more organ systems that require ongoing care from three or more specialists. The awardee specifies eligibility criteria for two program models on the basis of prior health service utilization. Eligibility criteria for the intensive model aim to identify patients who use tertiary centers at very high rates, including those who had (1) two or more hospitalizations totaling 10 or more hospital days, or (2) 20 or more outpatient clinic visits within a 12-month period. Eligibility criteria for the ambulatory model target children who use tertiary care centers at a moderately high rate and those who had (1) at least one hospitalization of 5 or more hospital days or (2) 10 or more outpatient clinic visits within a 12-month period. The hospitals do not vary the interventions provided to participants based on program model, so the evaluation considers participants in either model to be part of the treatment group.

[12] The presence of each health condition is defined as any diagnosis classified under the condition category by the PMCA being indicated in more than one claim.

Table III.2. Baseline-year demographic characteristics of Medicaid FFS and managed care beneficiaries enrolled in the Wisconsin Department of Health Services' program through December 31, 2015

	All enrollees (N = 203)	
	Number	Percentage
Age as of enrollment date		
0–2	102	50.2%
3–5	33	16.3%
6–10	31	15.3%
11–13	16	7.9%
14–17	21	10.3%
Gender		
Female	94	46.3%
Male	109	53.7%
Race/ethnicity		
White	49	24.1%
Black	14	6.9%
Other	15	7.4%
Unknown	125	61.6%
Geographic area		
Milwaukee metropolitan area	65	32.0%
Other metropolitan areas	80	39.4%
Rural areas	57	28.1%
Medicaid benefit plan(s)[a]		
SSI	158	77.8%
Foster care	14	6.9%
Other	45	22.2%
Managed care enrollment		
Any managed care enrollment	19	9.4%
No managed care enrollment	184	90.6%
Third-party insurance[b]		
Any TPL payment	83	40.9%
No TPL payments	120	59.1%
Medical complexity status[c]		
Non-chronic	8	3.9%
Non-complex chronic	4	2.0%
Complex chronic	189	93.1%

Table III.2 *(continued)*

	All enrollees (N = 203)	
	Number	Percentage
Body system(s) affected[d]		
Cardiac	92	45.3%
Craniofacial	4	2.0%
Dermatological	5	2.5%
Endocrinological	24	11.8%
Gastrointestinal	100	49.3%
Genetic	71	35.0%
Genitourinary	63	31.0%
Hematological	10	4.9%
Immunological	18	8.9%
Mental Health	92	45.3%
Metabolic	20	9.9%
Musculoskeletal	88	43.3%
Neurological	127	62.6%
Pulmonary-Respiratory	94	46.3%
Renal	14	6.9%
Ophthalmological	46	22.7%
Otologic	2	1.0%
Otolaryngological	13	6.4%

Note: The baseline period is defined as one year (365 days) prior to each individual beneficiary's program enrollment date. Medicaid enrollees must have at least 90 days eligibility in the baseline period and on the date of program enrollment date to be included in the eligible sample. All beneficiary characteristics are measured in the last month of the baseline period unless otherwise indicated.

[a]Beneficiaries may be eligible for more than one benefit plan; as a result, percentages may sum to more than 100.

[b]Identified in the beneficiary's one-year baseline period.

[c]Identified by using the Pediatric Medical Complexity Algorithm (PMCA) and in the three years prior to the beneficiary's program start date.

[d]Children may have more than one body system affected; as a result, percentages may sum to more than 100.

SSI = Supplemental Security Income; TPL = third-party liability.

3. Baseline characteristics for Medicaid beneficiaries, by program site

About twice as many beneficiaries first enrolled in CHW's program during the analysis time frame compared with those who first enrolled in AFCH's program (135 versus 68, respectively) (Tables III.3 and III.4). This difference may be related to the fact that there is a higher population concentration in Milwaukee and its surrounding areas compared with that in Madison. On the other hand, the difference may be related to the fact that CHW's program is well-established, having been in place (in a somewhat more limited form) since 2002, whereas AFCH's program

is relatively new.[13] At the start of the cooperative agreement in September 2014, CHW received a steady flow of patient referrals, whereas AFCH spent much of the first year of the cooperative agreement building connections with, and raising awareness of the SNP among, primary care providers, specialists, and community programs to foster future referrals.

The baseline beneficiary characteristics differ somewhat from one hospital to the other. Beneficiaries enrolled at CHW are younger, more likely to reside in a metropolitan area, and more likely to receive SSI benefits. They are less likely to have third-party insurance as the primary payer and Medicaid as the secondary payer, compared with those enrolled at AFCH. Specifically, more than half (60 percent) of beneficiaries enrolled at CHW are 2 or younger, but just 31 percent of those enrolled at AFCH are within this age range (Tables III.3 and III.4). It is likely that this difference reflects the fact that in the enhanced SNP's first year of operations, CHW actively recruited infants from the NICU. AFCH accepted infants from the NICU but did not actively recruit due to a well-established NICU follow-up program. Most beneficiaries enrolled at CHW (81 percent) reside in Milwaukee or in other metropolitan areas, and the remaining 19 percent reside in rural areas. Of those enrolled at AFCH, only about half (51 percent) reside in metropolitan areas, and the other half reside in rural areas. A somewhat higher percentage of beneficiaries enrolled at CHW (85 percent) are SSI recipients, compared with those enrolled at AFCH (63 percent). A lower percentage of beneficiaries enrolled at CHW (36 percent) had third-party insurance as the primary payer and Medicaid as the secondary payer, compared with those enrolled at AFCH (50 percent).

Despite these differences, beneficiaries at the two hospital sites were fairly similar at baseline in terms of rates of medical complexity and common diagnoses. Ninety-five percent of beneficiaries at CHW were classified as having complex chronic health conditions, and 90 percent of beneficiaries at AFCH were classified as such (Tables III.3 and III.4). Four of the top five most prevalent diagnoses were common to beneficiaries enrolled at both program sites. Neurological diagnoses were most common (59 percent of beneficiaries at CHW and 71 percent of beneficiaries at AFCH had a neurological diagnosis according to claims data). At both sites, other diagnoses among the top five included gastrointestinal diagnoses (46 percent of beneficiaries at CHW and 56 percent at AFCH), pulmonary-respiratory diagnoses (46 percent of beneficiaries at CHW and 47 percent at AFCH), and mental health diagnoses (43 percent at CHW and 50 percent at AFCH). Cardiac diagnoses were somewhat more prevalent among beneficiaries at CHW (52 percent versus 32 percent at AFCH), and musculoskeletal diagnoses were more prevalent at AFCH (54 percent versus 38 percent at CHW).

[13] AFCH launched the first iteration of a similar program in March 2014.

Table III.3. Baseline-year demographic characteristics of Medicaid FFS and managed care beneficiaries enrolled in the Wisconsin Department of Health Services' program through December 31, 2015, by program site: CHW

	All enrollees (N = 135)	
	Number	Percentage
Age as of enrollment date		
0–2	81	60.0%
3–5	19	14.1%
6–10	15	11.1%
11–17	20	14.8%
Gender		
Female	64	47.4%
Male	71	52.6%
Race/ethnicity		
White	29	21.5%
Black	11	8.1%
Other	11	8.1%
Unknown	84	62.2%
Geographic area		
Milwaukee metropolitan area	64	47.4%
Other metropolitan areas	46	34.1%
Rural areas	25	18.5%
Enrollment in other public benefits[a]		
SSI	115	85.2%
Other	28	20.7%
Managed care enrollment		
Any managed care enrollment	18	13.3%
No managed care enrollment	117	86.7%
Third-party insurance[b]		
Any TPL payment	49	36.3%
No TPL payments	86	63.7%
Medical complexity status[c]		
Non-chronic	5	3.7%
Non-complex chronic	1	0.7%
Complex chronic	128	94.8%
Body system(s) affected[d]		
Cardiac	70	51.9%
Craniofacial	2	1.5%
Dermatological	3	2.2%
Endocrinological	13	9.6%
Gastrointestinal	62	45.9%
Genetic	51	37.8%
Genitourinary	28	20.7%

Table III.3 *(continued)*

	All enrollees (N = 135)	
	Number	Percentage
Hematological	8	5.9%
Immunological	14	10.4%
Mental Health	58	43.0%
Metabolic	12	8.9%
Musculoskeletal	51	37.8%
Neurological	79	58.5%
Pulmonary-Respiratory	62	45.9%
Renal	6	4.4%
Ophthalmological	29	21.5%
Otologic	2	1.5%
Otolaryngological	8	5.9%

Note: The baseline period is defined as one year (365 days) prior to each individual beneficiary's program enrollment date. Medicaid enrollees must have at least 90 days eligibility in the baseline period and on the date of program enrollment date to be included in the eligible sample. All beneficiary characteristics are measured in the last month of the baseline period unless otherwise indicated.

[a]Beneficiaries may be eligible for more than one public benefit; as a result, percentages may sum to more than 100.

[b]Identified in the beneficiary's one-year baseline period.

[c]Identified by using the Pediatric Medical Complexity Algorithm (PMCA) and in the three years prior to the beneficiary's program start date.

[d]Children may have more than one body system affected; as a result, percentages may sum to more than 100.

SSI = Supplemental Security Income; TPL = third-party liability.

Table III.4. Baseline-year demographic characteristics of Medicaid FFS and managed care beneficiaries enrolled in the Wisconsin Department of Health Services' program through December 31, 2015, by program site: AFCH

	All enrollees (N = 68)	
	Number	Percentage
Age as of enrollment date		
0–2	21	30.9%
3–5	14	20.6%
6–10	16	23.5%
11–17	17	25.0%
Gender		
Female	30	44.1%
Male	38	55.9%
Race/ethnicity		
White	Not available[a]	Not available
Black	Not available	Not available
Other	Not available	Not available
Unknown	Not available	Not available

Table III.4 *(continued)*

	All enrollees (N = 68)	
	Number	Percentage
Geographic area		
Metropolitan areas	35	51.5%
Rural areas	32	47.1%
Enrollment in other public benefits[b]		
SSI	43	63.2%
Other	31	45.6%
Managed care enrollment		
Any managed care enrollment	Not available[a]	Not available
No managed care enrollment	Not available	Not available
Third-party insurance[c]		
Any TPL payment	34	50.0%
No TPL payments	34	50.0%
Medical complexity status[d]		
Non-chronic	3	4.4%
Non-complex chronic	3	4.4%
Complex chronic	61	89.7%
Body system(s) affected[e]		
Cardiac	22	32.4%
Craniofacial	2	2.9%
Dermatological	2	2.9%
Endocrinological	11	16.2%
Gastrointestinal	38	55.9%
Genetic	20	29.4%
Genitourinary	35	51.5%
Hematological	2	2.9%
Immunological	4	5.9%
Mental Health	34	50.0%
Metabolic	8	11.8%
Musculoskeletal	37	54.4%
Neurological	48	70.6%
Pulmonary-Respiratory	32	47.1%
Renal	8	11.8%
Ophthalmological	17	25.0%
Otologic	0	0.0%
Otolaryngological	5	7.4%

Note: The baseline period is defined as one year (365 days) prior to each individual beneficiary's program enrollment date. Medicaid enrollees must have at least 90 days eligibility in the baseline period and on the

Table III.4 *(continued)*

date of program enrollment date to be included in the eligible sample. All beneficiary characteristics are measured in the last month of the baseline period unless otherwise indicated.

[a]Number of beneficiaries in each category is too small to report.

[b]Beneficiaries may be eligible for more than one public benefit; as a result, percentages may sum to more than 100.

[c]Identified in the beneficiary's one-year baseline period.

[d]Identified by using the Pediatric Medical Complexity Algorithm (PMCA) and in the three years prior to the beneficiary's program start date.

[e]Children may have more than one body system affected; as a result, percentages may sum to more than 100.

SSI = Supplemental Security Income; TPL = third-party liability.

C. Identifying a comparison group

To identify a comparison group, we continue to explore methods for replicating the awardee's program eligibility criteria. Although we will be limited in our ability to replicate the clinical decision making that influences whether patients are enrolled in the enhanced SNP, we expect to be able to identify CMC based on the presence of three more condition categories flagged by the PMCA, coupled with information from claims data indicating children seen by three or more specialists. Furthermore, we also expect to be able to replicate the awardee's utilization-based medical fragility criteria by using information in claims data. Given the qualitative finding that many children are referred to the enhanced SNP following a major medical event (though it is not a program eligibility requirement), we will need to further explore how to replicate this characteristic when we select a comparison group.

Based on an analysis of the beneficiaries' zip codes, the enhanced SNP is not drawing Medicaid beneficiaries residing in northwest Wisconsin, which confirms that we may be able to draw a comparison group of beneficiaries from this area. One caveat is that beneficiaries with medical complexity or fragility in this area may travel to Minneapolis, the closest major metropolitan area, for tertiary care. However, the awardee confirmed that the timeliness and completeness of claims data for services received out of state should be comparable to that for in-state services, so we do not anticipate that this will pose a problem for the utilization and expenditures analysis.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

The Wisconsin Department of Health Services, a state Medicaid agency, developed a payment approach for SNP based on FFS reimbursement for targeted case management through an SPA. The Wisconsin Department of Health Services planned to submit the proposed SPA before the end of 2017 with rates effective retroactively to September 1, 2017.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor, previous narratives and payment model reports submitted to the implementation and monitoring contractor, and a semi-structured interview that we conducted with the awardee leaders who are most familiar with the payment model.

B. Description of the payment model

The state Medicaid agency planned to continue supporting the SNP based on FFS reimbursement for targeted case management through an SPA. The case management payments will cover program services not currently reimbursed under traditional Medicaid FFS and will include two procedure codes. Enrollment (G0506) will provide a one-time payment for comprehensive assessment and completion of a patient care plan. Ongoing care coordination (T2023) will provide a monthly capitated payment for referral, monitoring, and follow-up activities. Eligibility for the monthly care coordination payment requires a member of the child's care team to have at least one reciprocal contact with the child or child's family or on behalf of the child with another provider during the month. The one-time enrollment payment is approximately $1,100, and the ongoing care coordination rate is $450 per participant per month. Providers cannot bill for enrollment and ongoing care coordination in the same month. The Wisconsin Department of Health Services payments will cover the cost of the program for Medicaid participants who account for approximately 85 percent of total participants in the SNP. Hospitals will not receive reimbursement through this plan to deliver services to participants without Medicaid coverage.

C. Status of developing the payment model

The Wisconsin Department of Health Services planned to submit the proposed SPA before the end of 2017 with rates effective retroactively to September 1, 2017, if approved by CMS. The awardee engaged the state actuary to determine a per beneficiary per month program cost for SNP members to inform payment model development. The two participating sites completed a three-month intensive time study from January through March 2017. The awardee used data from the time study combined with monthly invoices, monthly enrollment files, and estimated enrollment at full program capacity to analyze program costs and define payment rates for case management. In addition, the Wisconsin Department of Health Services had begun working on an operating agreement between the department and its hospital partners to establish quality metrics as part of the payment model. The awardee initially will focus on process measures and later move toward including outcome measures similar to those of interest during the cooperative agreement such as inpatient days and ED visits. The awardee planned to also develop a plan to

engage private payers after they have finalized and implemented the payments for Medicaid case management.

According to the awardee, the modest costs of SNP with the potential for significant savings warranted continuing the program at least in the short term. The Wisconsin Department of Health Services has observed a significant decrease in participant Medicaid costs two years after SNP enrollment compared to two years prior. However, program leaders also acknowledged that cost savings could be attributable to the program or could reflect reduced medical costs following an acute episode that prompted a Medicaid member to enroll in SNP. Therefore, as more data become available over the next few years, program leaders plan to review program sustainability. Once they have more information about the program's return on investment, they will decide whether to continue the current case management model or consider using an alternative payment model based on shared savings. Finally, the awardee reported that, at current staffing levels, the hospitals have some capacity to increase enrollment, which would reduce or reverse any gap between program costs and Medicaid reimbursement.

D. Factors associated with development of payment model

The Wisconsin Department of Health Services' status as the state Medicaid agency and strong working relationship with CHW and AFCH during the cooperative agreement period facilitated development of the payment model. Both partner hospitals and their providers actively engaged in the process, including completing the three-month time study to inform model development and providing additional data on program costs to support detailed cost analyses. A program leader also noted the contribution of CMMI in fostering an environment of collaboration and offering feedback to assist in developing the SNP payment model.

The biggest challenges the Wisconsin Department of Health Services faced in its model development work included lack of a robust program evaluation and a compressed time frame to collect data to assess savings. One program leader explained, "Looking at savings over time takes time, and a three-year award period is really two years of data. To base a payment model on two years of data is tough. That's why with this targeted case management payment, we're essentially extending that sort of review period for another two years."

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

The Wisconsin Department of Health Services and its hospital partners, CHW and AFCH, decided to use their own, current staff and other resources to sustain the SNP while they awaited approval for their proposed payment model (as discussed Chapter IV). The awardee received a 12-month NCE but intended to use the additional time to complete program requirements and advance the evaluation, while supporting the program with hospital resources and the proposed payment model, if CMS approved the SPA. CHW had broadened eligibility for the program, and the payment model would promote scaling the program to additional children's hospitals with sufficient capabilities to implement the full model. Because the Medicaid payment model does not cover program costs for commercially insured participants in the program, the awardee also aimed to facilitate hospital efforts to engage commercial payers. The awardee had not been involved in helping other providers replicate the program elsewhere.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

In Year 2, the Wisconsin Department of Health Services had begun actively pursuing strategies to sustain the SNP. The awardee considered Section 1945 Health Home and Section

1905 Targeted Case Management (TCM) for a Medicaid SPA for the program. For longer term sustainability, the awardee was planning to adopt an alternative payment model based on hospital programs' accountability for participant outcomes. In addition, the Wisconsin Department of Health Services was planning to scale the program to engage more hospitals, in part through its SPA application. At that time, others had not replicated the program nor was the awardee actively working to foster replication.

C. Implementing the SSR plan: progress and changes

Sustainability. As described in Chapter IV, in Year 3, the Wisconsin Department of Health Services was awaiting approval from CMS for its SPA, which included two billing codes for the initial assessment and ongoing case management functions of the program, and was considering pursuing an alternative payment arrangement with shared savings down the road. Because of the program's potential to save money, at least one of the hospitals (CHW) expected to sustain the program with the same number of staff using its own funds while awaiting the SPA decision, which, if approved, would be retroactive to September 2017.

However, the Wisconsin Department of Health Services shared some concerns about the long-term sustainability of program. Specifically, the program accepts beneficiaries regardless of insurance type, and Medicaid cannot subsidize services for non-Medicaid participants. Through its support for the SNP, including demonstrating the value of the program and developing actuarially sound payment rates, the awardee aimed to facilitate engagement of commercial payers by hospital partners to support the program.

Scalability. The SNP represented an expanded version of an existing program at CHW. Under the cooperative agreement, CHW scaled its existing program to an expanded population by broadening its eligibility criteria. In addition, the cooperative agreement enabled the SNP to be scaled to AFCH, which had not fully established an SNP before the cooperative agreement.

Neither CHW nor AFCH reported plans to further scale the program. CHW leadership reported that it had no plans to further expand its eligibility criteria, but instead wanted to maintain focus on the sickest, most challenging children likely to use inpatient services. One interview respondent said that, in fact, implementing the new payment model might necessitate tightening the eligibility. AFCH leaders expressed similar concerns and did not want to expand the program beyond the population for which it believes it can provide the most benefit.

The Medicaid program planned to limit the type of providers that can bill for the new care coordination codes (if approved), which would restrict the extent of the program's scalability. To ensure that participating hospitals offer the same level and quality of care as the existing program, only free-standing children's hospitals with access to a specified range of subspecialists would be allowed to use the billing codes. A new children's hospital opening in the state had not approached the awardee about joining the program. The awardee planned to wait and see whether the hospital had sufficient capability before considering scaling the program.

Replicability. The program's replication activities were limited to disseminating information about the SNP. The hospitals presented at lectures and conferences in other states and focused on publishing the results of the program in academic journals.

D. Factors associated with progress toward implementing the SSR plan

The SNP had strong leadership support from the Wisconsin Department of Health Services, which should help with sustaining and, potentially, scaling the program. The awardee reported it had demonstrated to its leaders that the program "seals a void" in caring for the target population, and that the program has "excellent allies" among specialty physicians who value the program. As a longstanding program, the program at CHW enjoyed strong support from hospital administration, which provided gap funding to cover the costs the program could not cover on its own. At AFCH, the program was newer and, although hospital leaders were in favor of the program, program leaders did not expect the organization to provide such gap funding. However, the awardee was concerned that hospital leaders' commitment to the program could wane if the hospitals needed to absorb more costs over time.

In addition to the challenges discussed in the Payment Model chapter, the awardee faced the challenge of an uncertain state Medicaid budget and expected it would need to convince Medicaid leaders that the payments were important for both fiscal and quality of care perspectives. Although it did not yet have complete information, the awardee was optimistic it could demonstrate that the program will be cost-neutral to the state and will provide a higher level of care to participants than they would otherwise receive.

This page has been left blank for double-sided copying.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, and the awardee received a no-cost extension, it will not enroll new beneficiaries. Therefore, the implementation evaluation team has concluded primary data collection unless CMS asks us to follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans. We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor, to determine whether there is information that will provide additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experience that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

During the coming year, we will select a comparison group for participants in Wisconsin Department of Health Service's SNP. Using propensity score matching, we will select beneficiaries whose characteristics closely resemble those of the program participants. Limiting the comparison group to Medicaid beneficiaries whose observed characteristics closely match those of the beneficiaries in the treatment group will tend to reduce the unobserved differences between the two groups if those characteristics are correlated with matching variables. We will convey the results of the match, including a balance chart, to CMS as well as impact findings as they become available.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

MATHEMATICA
Policy Research

REPORT

HCIA Round Two Evaluation: Yale University

April 18, 2018

Sonya Streeter (Mathematica Policy Research)
Rumin Sarwar (Mathematica Policy Research)
Eva Chang (RTI International)

Submitted to:

Center for Medicare & Medicaid Innovation
Rapid Cycle Evaluation Group
7500 Security Boulevard, Mailstop 06-05
Baltimore, MD 21244

COR and Evaluation Co-leader: Jean M. Gaines
Evaluation Co-leader: Patricia Markovich
Contract Number: CMMI-500-2014-00034I

Submitted by:

Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Randall Brown
Reference Number: 50082

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURE

I. INTRODUCTION

A. Background of HCIA R2 and purpose of the awardee narratives

On September 1, 2014, the Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) awarded the second round of cooperative agreements, known as Round Two of the Health Care Innovation Awards (HCIA R2), to 39 organizations.[1] The 39 awardees used these three-year cooperative agreements, which ended on August 31, 2017, to implement their proposed innovative service delivery models, to design and test new payment models for improving health and the quality of care, and to lower the cost of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) beneficiaries.[2]

The federal evaluation of HCIA R2 has twin goals. The first is to describe the awardees' innovative programs, assess the extent to which they have been successful in implementing their programs, and identify the factors that facilitated or impeded their implementation progress (referred to as the implementation evaluation). The second goal is to estimate the impact of the programs on costs, hospital admissions, unplanned readmissions, and emergency department (ED) visits—the target outcomes of the impact evaluation.

In this year's awardee narratives, the implementation evaluation focuses on determining whether awardees were able to implement their programs in a way that was likely to reduce costs and improve health outcomes.[3] In addition to updating what programs were doing in the last year of their three-year cooperative agreements, the implementation evaluation assesses (1) whether awardees achieved implementation effectiveness, defined as meeting their enrollment and service delivery goals, and (2) whether programs had their intended effect on the delivery of care based on the perceptions of participating clinicians and other intervention staff (Chapter II).

The impact evaluation may also include (depending on the evaluability of each program and the status of the impact analysis for programs that are evaluable) the results of our efforts to identify a matched comparison group and the preliminary findings from our analysis of the program's impact on the core outcomes (Chapter III). The narratives also describe the awardees' progress in developing their payment models (Chapter IV) and in updating their plans for sustaining, scaling, or replicating their programs now that the initial three-year cooperative agreements have ended (Chapter V). Last, we describe the remaining activities in the final phase of the federal evaluation (Chapter VI).

B. Overview of the awardee's program

Yale University used HCIA R2 funding to implement the Paramedic Referrals for Increased Independence and Decreased Disability in the Elderly (PRIDE) program. PRIDE used a

[1] George Washington University voluntarily terminated its cooperative agreement with CMMI, effective September 1, 2016, and is not included in this report.

[2] Yale University received a 12-month no-cost extension through August 31, 2018. Yale University will enroll participants through February 28, 2018, and provide program services through May 31, 2018.

[3] The second annual report, released in August 2017, is available at https://downloads.cms.gov/files/cmmi/hcia2-yrtwoannualrpt.pdf.

community-based, care management approach to improve the health of elders and others with impaired mobility in an effort to prevent injuries from a fall and associated hospitalizations. The awardee observed that individuals who have fallen or fear falling may not have a primary care physician (PCP) to address their underlying health issues or notified their PCP about a previous fall experience. PRIDE sought to engage these individuals in their home to identify fall risks, address underlying medical conditions that lead to fall risk, provide referrals to relevant community resources, and implement fall prevention strategies.

PRIDE services included in-home assessments related to fall risk, in-home preventive care (such as arranging installation of durable medical equipment or conducting medication reconciliation), and linkage to primary care (such as scheduling a PCP appointment or arranging transportation services to/from one PCP visit). PRIDE paramedics and nurses from partnering visiting nurse agencies (VNAs) provided these services. Program leaders expect that PRIDE services will reduce falls and thus contribute to reductions in preventable ED visits, hospitalizations, and 911 calls. Ultimately, successful reductions in fall rates should reduce mortality and morbidity in the target population.

Table I.1. HCIA R2 program characteristics at a glance

Program characteristic	Description
Purpose	In the PRIDE program, Yale University serves participants who have fallen or are at risk of falling and provides them with in-home interventions and increased linkages to PCPs. Yale aims to reduce falls and other medical emergencies that contribute to preventable ED visits, hospitalizations, and 911 calls.
Major innovation	Employ paramedics and VNA nurses to visit participants in their homes in order to identify their medical and social needs and connect them to appropriate resources.
Program components	Care management
Target population	Individuals in the greater New Haven area of CT who have fallen or who are at risk of falling. Yale does not restrict enrollment to elderly individuals, but anticipates that a majority of enrollees will be over age 65.
Theory of change/ theory of action	If a paramedic and VNA nurse visit participants in their homes to perform fall risk assessments, deliver preventive care, and link participants to primary care, the number of falls and other medical emergencies will be reduced.
Payment model	Capitated payment for care management/coordination services
Award amount	$7,159,976
Effective launch date	March 25, 2015
Program setting	Participants' homes
Market area	Urban and suburban
Market location	Greater New Haven, CT
Target outcomes	Reduction in lift assist calls, mortality, ED visits, hospital admissions, and total expenditures in the target population.

C. Summary of the main findings

Implementation evaluation. Based on a review of the information available to us, we have concluded that the awardee was partially successful in implementing its program by the end of the initial three-year cooperative agreement. We based this conclusion on five factors. First, the awardee enrolled 4,288 participants—116 percent of its three-year projection—by the end of the initial cooperative agreement. After experiencing low enrollment during the first program year, the awardee overcame challenges to enrollment by expanding the eligibility criteria and identifying additional recruitment opportunities. Second, the awardee provided PRIDE services as intended, although it struggled to meet demand after experiencing growth in enrollment at the end of the first program year and before hiring and staffing was complete. Third, the awardee overcame early recruitment and hiring challenges and implemented creative staffing model changes in the second and third program years—which resulted in having a sufficient number of well-trained staff to deliver program services as planned during the third program year. Fourth, after experiencing low uptake of program services among enrollees during the first program year, the awardee implemented strategies that improved participant engagement and the number of participants accepting program services during the final two years of the cooperative agreement period. Finally, program leaders and staff felt that the program had a positive effect on the care provided and outcomes achieved by program participants.

Impact evaluation. Due to the lack of a strong comparison group, we do not anticipate being able to conduct a rigorous impact analysis for Yale University. However, the impact evaluation in this narrative does include a description of the baseline characteristics of program participants.

Payment model. Yale University designed a payment model in which the regional emergency medical services (EMS) medical director receives prospective payments for a geographical region. The medical director serves as an intermediary and reimburses paramedics, visiting nurses, and scheduled transportation providers for providing specific services, including an initial in-home assessment by the responding paramedic, a two-hour initial home visit by a visiting nurse, and scheduled round-trip transportation to the participant's PCP.

Sustainability plans. Yale University plans to use the no-cost extension period to develop a sustainability plan and transition the program from an academic setting to the public/private sector. Specifically, Yale University will continue enrolling and providing program services, engaging payers after collecting and analyzing data to inform payment model development, and considering regulatory options to address the paramedics' scope of practice requirements. Yale University did not report plans to scale or replicate its program.

This page has been left blank for double-sided copying.

II. FINDINGS FROM THE IMPLEMENTATION EVALUATION

A. Data and methods for assessing implementation effectiveness

The findings presented in this chapter are based on the evaluation team's analyses of three data sources. The first was the awardee's self-reported information submitted quarterly to the implementation and monitoring contractor from September 2014 through August 2017. This information included the program's operating plans along with the awardee's self-measurement and monitoring reports, program narratives, progress reports, and supplemental materials. The second data source was qualitative information collected annually toward the end of each of the three years of the cooperative agreement through in-person and virtual site visits. We used semi-structured protocols tailored to the unique features of each program to interview awardee administrators and frontline staff responsible for implementing the program at selected sites. The third source included responses to two surveys—one of clinicians and one of non-clinician staff—on their perceptions of the effect of the program on the delivery of care. The non-clinician staff survey was fielded from July to October 2016 with a sample of 44 respondents, and achieved a response rate of 50 percent. We did not weight the survey samples to adjust for nonrespondents. For survey items to which fewer than 11 individuals responded, we describe the findings in qualitative terms to avoid identifying respondents.

To assess implementation effectiveness, we considered each awardee's enrollment and its success in implementing the service delivery model. We assessed enrollment based on how close the awardee came to meeting its enrollment goal. We assessed the implementation of the service delivery model on the basis of four domains: (1) delivery of intervention services, (2) staffing and training, (3) recruitment and engagement of providers, and (4) engagement of participants (Table II.1). Our primary objective was to determine whether the program—as it was implemented—was likely to lead to lower costs or improved outcomes.

Table II.1. Criteria used to assess implementation effectiveness

Component	Domain	Criteria
A. Program enrollment		• Did the awardee meet, or nearly meet, its enrollment goal in a timely manner?[a]
B. Service delivery	1. Delivery of intervention services	• Did the awardee deliver intervention services consistent with the planned mode, frequency, intensity, duration, and quality? • If the awardee changed its service delivery protocol, did deviations in services affect the ability to realize desired outcomes?
	2. Staffing and training	• Did the awardee recruit and hire qualified program staff in a timely manner and retain most of them throughout the cooperative agreement? • Did the awardee provide relevant and effective training in its service delivery protocol (if applicable)?
	3. Recruitment and engagement of providers	• Did the awardee meet, or nearly meet, its provider recruitment goals? • Did the awardee engage participating clinician and non-clinician staff in the delivery of intervention services in a timely manner and in a meaningful, sustained way (if applicable)?
	4. Engagement of program participants	• Did the awardee engage participants in a timely manner and in a meaningful way? • Did the awardee retain most of participants for the full period of enrollment (as applicable)?

[a] Assessment of awardees meeting various percentages of their three-year enrollment projections by the end of the 12th program quarter: 90 percent or more = achieved enrollment effectiveness; between 65 and 90 percent = partially achieved enrollment effectiveness; and less than 65 percent = did not achieve enrollment effectiveness.

B. Assessment of implementation effectiveness

Based on the evidence available to us, we conclude that the awardee was partially successful in implementing its program. We based this conclusion on the preponderance of evidence as it relates to the two main program components: enrollment and implementation of the service delivery model. Separately for each component, we (1) describe the awardee's implementation strategy and highlight significant changes made during the three-year cooperative agreement, (2) summarize the evidence on which we base our assessment, and (3) describe the factors that helped or hindered the achievement of program goals.

1. Identification, recruitment, and enrollment of the target population

a. Description of and changes to recruitment and enrollment strategies

Yale University actively enrolled direct participants after obtaining the participants' consent. Participants remained enrolled in PRIDE until they had either completed or declined both the paramedic and VNA nurse visit.

Yale University's approach to participant identification and recruitment evolved throughout the cooperative agreement period, after recognizing that the program's initial focus was too narrow and contributed to low enrollment. Originally, the awardee only recruited patients who had fallen and called 911 for a "lift assist" but were determined to be uninjured and not transported for further medical attention. Toward the beginning of the second program year, Yale University expanded the participant criteria to include individuals who expressed concern about

a fall, even if they had never experienced one. In addition, Yale University recruited patients who might benefit from the program who were admitted to the hospital from the ED and then discharged home. To recruit patients from the expanded target population, Yale University enrolled participants through three identification and recruitment mechanisms:

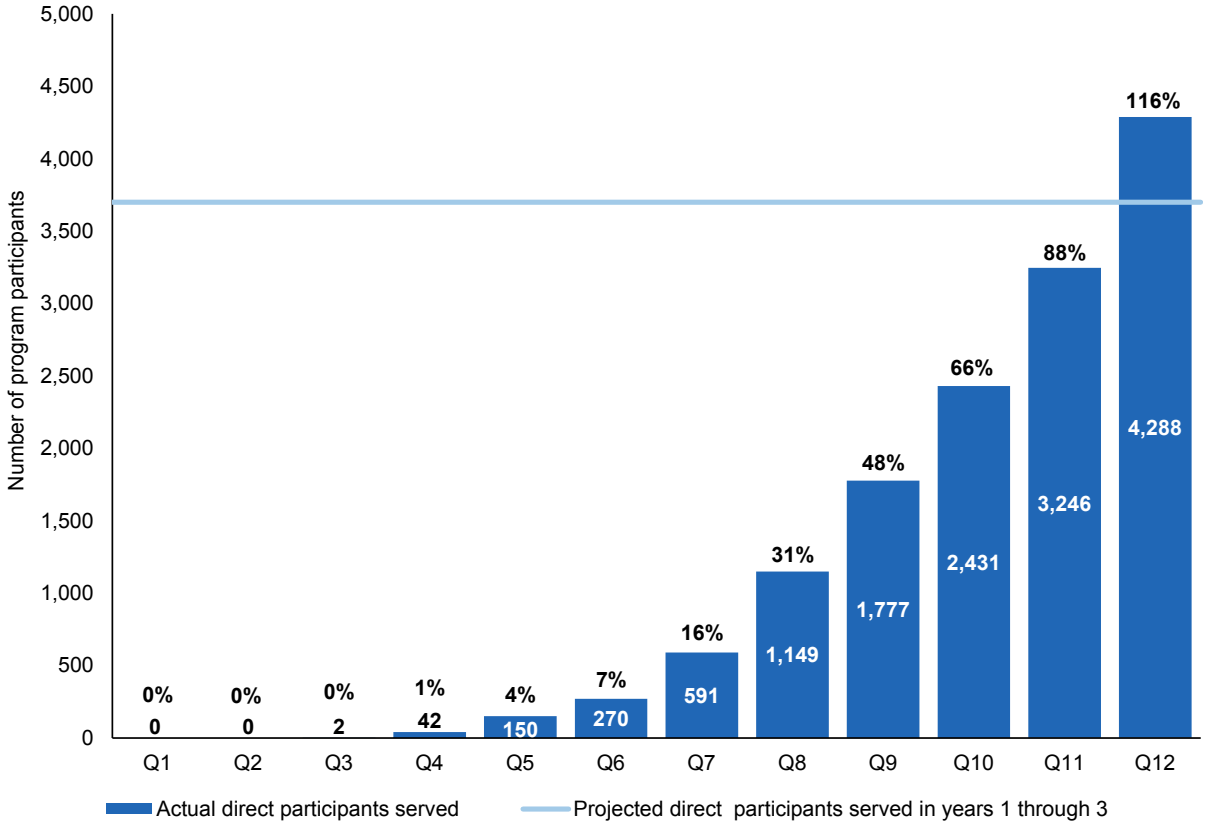
1. **After a 911 call.** During a 911 visit, a responding EMS provider identified a patient who appeared at risk for falls, described the PRIDE program, and provided an informational brochure. A PRIDE paramedic then contacted the patient at home shortly after the 911 visit to provide more information about the program and complete enrollment.
2. **While in the ED.** Yale University hired PRIDE researchers to speak with potential participants who sought medical care in the ED. The researchers described the PRIDE program to ED patients and their caregivers, either to complete enrollment at that time or to collect contact information so that PRIDE paramedics could visit them at home to complete the enrollment. Based on the success of this enrollment strategy, Yale University expanded recruitment and enrollment efforts to a second ED in the 6th program quarter and to a third ED in the 12th program quarter.
3. **Based on self-referrals.** Toward the end of the first program year, individuals or their caregivers enrolled after learning about PRIDE from a presentation at a community event (for example, a town hall or senior center meeting), through social media (for example, the PRIDE website or Google+ site), or via the PRIDE brochure. Individuals also enrolled during the community event or called the PRIDE central office to enroll. Based on the success of this enrollment strategy, Yale University expanded its catchment area for community recruitment events during the third program year to respond to requests by local housing authorities to serve participants in their areas.

b. Evidence of enrollment effectiveness

Yale University achieved partial enrollment effectiveness. Overall, the awardee reported that it enrolled 4,288 participants from March 2015 (when it launched its program) through August 2017, about 116 percent of its 3,700 three-year projected participants (Figure II.1). Early difficulties with enrollment resulted in the awardee enrolling 42 participants as of the first program year—11 percent of the 400 participants that the awardee had expected to enroll by the end of the first program year. In response to low enrollment numbers, Yale University reset the three-year enrollment projection in the ninth program quarter and reduced the original enrollment target (4,800) by 23 percent. By the end of the third program year, Yale University met 89 percent of its original enrollment target.

According to the awardee's self-monitoring data, by the end of the cooperative agreement period, ED visits, and self-referrals each made up 49 percent of total enrollment, making them the most successful of Yale University's three identification, recruitment, and enrollment mechanisms (2,099 and 2,086 participants, respectively). The other 2 percent of enrollees came through referrals of patients who called 911 for a lift assist (103). Since the awardee did not submit information about the characteristics of the participants who were enrolled, we are unable to assess the awardee's success in serving participants who matched its target population.

Figure II.1. Projected versus actual cumulative direct participants served through Year 3, as of August 31, 2017



Source: Enrollment data from the implementation and monitoring contractor, program quarters 1 through 12 (September 2014–August 2017).

Note: Projected direct participants served reflects the cumulative and unique number of individuals whom the awardee expected to serve in the program through August 2017. Direct program participants refers to the total number of unique participants who received a service directly paid for by HCIA R2 funding. The awardee lowered its direct participant enrollment projection from 4,800 at the beginning of the first program year to 3,000 at the beginning of Year 2. It then adjusted its direct participant enrollment projection to 3,700 at the beginning of Year 3.

c. Barriers and facilitators associated with enrollment effectiveness

Yale University's progress in meeting its three-year enrollment goals was influenced by several factors. First, building strong relationships with ED staff helped boost enrollment of ED patients. Second, the awardee implemented strategies to overcome community mistrust of the program and language barriers with non-English-speaking communities. Finally, although the awardee took steps to increase enrollment of participants after a 911 call, this recruitment strategy remained unsuccessful.

Program staff reported that strong relationships with the ED recruitment sites facilitated enrollment. Program leaders regularly communicated with ED management about PRIDE, who in turn communicated with ED staff to increase their understanding and awareness of the program. According to program staff, ED nurses and doctors who were knowledgeable and appreciative of the program were more likely to flag appropriate patients for enrollment or

communicate to PRIDE staff which patients should not be approached, for example, patients with signs of dementia or who were seeking treatment for a serious medical condition. ED management also allowed PRIDE staff to leave brochures in the ED waiting room; according to program staff, this appeared to increase potential participants' awareness of the program.

> "When I enroll patients, sometimes they already know about the program. They've...seen that poster at the senior center, or [their] friend does the program. It's not an alien concept to them... I think that makes them more willing to participate."
>
> —Program staff

Yale University's efforts to implement strategies that increased community awareness of PRIDE purportedly created a positive reputation for the program and built trust between the program staff and community, helping to overcome the public's initial mistrust of the program and facilitating enrollment efforts. The awardee became aware of the public's initial mistrust of the program after collecting feedback from potential participants during the first program year; they learned that the target population was suspicious of the PRIDE in-home intervention and concerned that the assessments would result in loss of independence. In addition, local housing authorities in New Haven initially resisted PRIDE recruitment. Program leaders believed this was due to the community's distrust of outsiders and tension between Yale University and city management. To address public misunderstanding about the program, Yale University initiated a social media campaign in the second program year that included billboards, brochures, and radio advertisements. Multiple program staff and leaders also mentioned that the program's branding efforts—such as using PRIDE logos on badges and uniforms—appeared to increase the community's trust in the program. The value of PRIDE branding was most notable for VNA nurses, who initially wore their VNA uniforms—making individuals worry that VNA nurses who worked for PRIDE would refer participants to hospice, one management staff explained. Program staff also said that reviewing potential participants' EMR data before approaching them allowed them to tailor the recruitment conversation to the patient's specific challenges and, thus, facilitated enrollment by helping some participants overcome initial mistrust of the program. By the third program year, the local housing authorities had also learned about the benefits of PRIDE from the marketing campaigns and word of mouth and proactively reached out to request PRIDE recruitment events for residents in their apartment buildings.

Yale University realized that additional steps were required to engage the large, non-English-speaking communities in the greater New Haven area. Awardee leaders believed that these communities included a high-need and underserved population that would benefit from PRIDE's in-home intervention. To improve communication with and enroll larger numbers of non-English-speaking patients, Yale University acquired translation services and hired multilingual staff for ED site recruitment activities and community events. Management and leadership staff reported that adding translation services supported the "increase in enrollment within the Hispanic community...even though [patients understand] both languages, they feel more comfortable in their language."

Yale University unsuccessfully attempted to increase the number of participants enrolled after a 911 call and identified multiple barriers that thwarted recruitment efforts through this mechanism. Early in the first program year, for example, the awardee encountered unexpected challenges acquiring timely 911 data to identify potential participants. In addition, program staff found that older adults often did not like being approached after a 911 lift assist call. In the second program year, the awardee offered first responders a $5 gift card to a local grocery store for every successful referral to PRIDE; this approach did not appear to substantially improve 911-based enrollment rates in the third year.

> "After a fall, patients are afraid or embarrassed. They just want to get back to what they were doing. They are not receptive at that time to sign up for something...They did not like the intrusion. It was like a sales call and they didn't trust it."
>
> —Program staff

2. Delivery of program services

a. Description of and changes to service delivery model

By delivering PRIDE program services, Yale University sought to improve the health of elders and others with impaired mobility in order to prevent injuries from a fall, using a community-based, care management approach. Through the work of PRIDE paramedics and VNA nurses, the awardee sought to identify participants' fall risks, address underlying medical conditions and home environment factors that could increase fall risks, improve linkages between participants and primary health care, and provide referrals to relevant community resources. The awardee also relied on multiple off-the-shelf, web-based information technology tools to operate the PRIDE program. In this section, we describe the key activities of the PRIDE service delivery model that the awardee envisioned would lead to desired outcomes.

A PRIDE paramedic completed the first in-home assessment, which considered multiple aspects of the participant's health status and residential safety. At the conclusion of the assessment, and if the participant agreed, the PRIDE paramedic scheduled an appointment for the participant with his or her PCP. If the participant desired transportation assistance, the paramedic arranged for transportation to and from one PCP visit. In the third program year, the awardee set a goal of completing the paramedic home visit within 14 days of enrollment; there was no defined time frame for completing the paramedic visit during the first two program years.

A PRIDE VNA nurse completed the second in-home visit, which involved a participant health assessment and preventive care measures (for example, appropriate configuration of a walker). The VNA nurse assessed the individual's overall health, with particular attention to the following: determining the need for ongoing nursing care, physical therapy, occupational therapy, or durable medical equipment; assessing for suspected neurological conditions; assessing mobility; and performing a medication review and reconciliation. The VNA nurse also acquainted the participant with the resources and services available within the community and how to access them, and educated the participant and household members on fall prevention strategies. The VNA nurse communicated the assessment findings to the participant's PCP, depending on the participant's assessment results. In the third program year, the awardee set a goal of completing the VNA home visit within 30 days of enrollment, but did not define a time frame for completing the VNA nurse visit during the first two program years. Also in the third program year, Yale University adapted the program to offer a combined paramedic and VNA

nurse visit if a participant required translation services or lived in a neighborhood perceived by staff to be unsafe.

Yale relied on multiple off-the-shelf, web-based information technology tools to operate the PRIDE program. Beginning in the first program year, intervention staff used Research Electronic Data Capture (REDCap) to record and review participant data, including demographic and contact information and assessment data. Over the three program years, Yale University relied more heavily on REDCap to track additional participant-related data and, most notably, schedule visits and services. In the second program year, Yale incorporated ShiftPlanning into operations to more efficiently align paramedic and VNA nurse availability with participants' desired times for home visits. Yale also adopted a MetroTaxi application in the third program year to request and monitor transportation assistance.

PRIDE grew out of the project leaders' previous experience implementing a program that worked with patients who called 911 for a lift assist. Earlier interviews indicated that the awardee leveraged knowledge gained from implementing the first program, but the two programs are substantially different in terms of the services provided, care teams, and intended goals.

b. Evidence of service delivery effectiveness

Yale University was partially successful in achieving service delivery effectiveness. The awardee was able to deliver program services as intended during home visits to enrolled participants, although it struggled to meet demand after experiencing growth in enrollment at the end of the first program year and before hiring and staffing was complete. The awardee successfully overcame early recruitment and hiring challenges by adapting its staffing model in the second and third program years, and successfully recruited, hired, and trained more program staff (paramedics and a dispatcher) and partnered with an additional VNA. In addition, adjustments to participant communication strategies and streamlining of program operations led to an increase in participants accepting home visits, although uptake rates of services that created or reestablished linkages to primary care remained low. We provide more information below.

Delivery of intervention services. Yale University was successful in offering and delivering PRIDE services to participants. Program leaders and staff reported that, as intended, program staff offered to schedule home visits for participants as soon as they enrolled. In cases where participants accepted and completed home visits, program leaders and staff reported that services were provided as outlined in the PRIDE service delivery model and that PRIDE staff consistently followed service delivery protocols. During interviews, program staff described conducting assessments, making appointments for participants with their PCPs and providing transportation services when needed, and identifying and resolving medical and environmental factors associated with fall risks.

Staffing and training. Yale University was partially successful in recruiting, hiring, and retaining program staff needed to deliver program services. Awardee leaders reported delays in hiring and onboarding paramedic staff at the start of the three-year cooperative agreement, which limited the program's capacity to provide services in both the first and second program years. Program leaders reported overcoming hiring challenges and maintaining an adequate and appropriate staffing level by the end of the third program year. They also reported that staff they

hired were high quality individuals, described by a program leader as "people [who] are personable, outgoing, who care about the community, and want to make a difference."

Yale University successfully trained program staff for their roles in the program. The awardee reported, for example, that as of August 2017, program staff had participated in 1,784 hours of training. According to respondents to the non-clinician staff survey, the most common types of training in which staff participated included asking a colleague for help (100 percent or 22 people), staff meetings (100 percent or 22 people), shadowing others (95 percent or 20 people), and self-study, including reading PRIDE protocols (86 percent or 19 people). Paramedics, for example, shadowed an ED case manager as part of their training. According to one interviewee, this allowed her to learn how to avoid defaulting to the "mentality of [a first responder] to rush in, do whatever they have to do, and rush out," and instead, listen carefully to participants to understand their challenges and the influence of those challenges on their well-being. In addition, PRIDE paramedics and VNA nurses received hands-on REDCap training, according to program staff and leaders. Almost all survey respondents agreed that the training taught them new skills that were important for their role with the program and helped improve their job performance (84 percent or 16 people).

Recruitment and engagement of providers. Yale University did not include recruitment or engagement of PCPs or other community service providers as a core component of the PRIDE service delivery model.

Engagement of program participants. Yale University was partially successful in engaging participants in the PRIDE program. During the first program year, the awardee struggled with engaging participants. According to self-monitoring data submitted by the awardee, 44 percent of all enrolled participants declined both the paramedic and VNA home visits in the first program year. In the next two years, the percentage of participants declining both home visits dropped to 25 percent and then to 16 percent, respectively. Yale University also experienced low uptake of services to create or reestablish linkages to primary care among participants who completed home visits. According to the awardee's self-monitoring data submitted in August 2017, approximately one in five participants accepted PRIDE paramedics' offers to schedule PCP appointments; only 3 percent of those who accepted this service requested transportation to a PCP appointment over the course of the three-year cooperative agreement.

Program staff were optimistic about the program's ability to engage participants, especially toward the end of the cooperative agreement, and reported receiving positive feedback from participants and other community members about PRIDE. In addition, all staff who completed the non-clinician survey at the beginning of the third program year agreed that the program had successfully engaged participants in the program (22 people). Almost half reported that resistance to the program by participants was not a barrier to achieving program goals (48 percent or 10 people), and just over half reported that participant engagement helped the program achieve its goals (59 percent or 13 people). According to survey respondents, the primary reason that participants may have declined participation in the program was due to not wanting to be part of a research study.

c. Barriers and facilitators associated with service delivery effectiveness

Several factors affected the awardee's ability to deliver program services as intended. Looking over the full three-year cooperative agreement, the awardee identified three primary facilitators: (1) effective staff communication and teamwork, (2) creative modifications to the PRIDE staffing model, and (3) use of REDCap to collect and communicate data. Program leaders and staff also described two primary barriers to service delivery that impacted uptake of program services: (1) participants' mistrust of PRIDE's VNA home visits and (2) challenges in engaging the target population. In addition, awardee leaders described Yale University's standardized hiring process as burdensome and time-consuming, resulting in delays in the first program year related to onboarding new PRIDE paramedics and recruiting new hires.

First, program staff and leaders at Yale University believed that strong communication and teamwork among program staff, including collaborative problem-solving and the sharing of best practices, facilitated service delivery. Program staff were encouraged to join all-staff quarterly meetings to discuss challenging cases, brainstorm solutions, and reflect on recent program data. Additionally, paramedics and ED recruitment staff held separate monthly meetings to share new insights about the program. In interviews, staff said they appreciated how these meetings solidified relationships with their peers and other types of program staff. Staff also valued the opportunity to discuss and update service delivery protocols (such as statement of work and standard operating procedures) to reflect new best practices.

> "I think the success [of PRIDE] comes down to a team effort, from the top all the way to the bottom. Everyone is pulling together to make this a success. I think that's been absolutely critical."
>
> —Program management staff

Second, the awardee's success in creatively adapting its staffing model in the second and third program years facilitated service delivery. During the second program year, for example, PRIDE hired a dispatcher in their central office to schedule home visits and conduct participant calls, among other tasks. According to program leaders and staff, having the dispatcher call participants to schedule, reschedule, and remind them of their appointments increased scheduling efficiency and decreased participant no-shows at home visits. The dispatcher also answered the phone and responded to questions by participants and program staff. During the third program year, the awardee further altered its staffing model, which previously had relied on shift-based and partner staff only, by hiring a full-time nurse and a full-time paramedic. Because the full-time staff had more time than the shift-based and partner staff, the awardee was able to deliver home visit services in a more timely fashion, thereby addressing concerns about the growing volume of pending visits.

Third, program staff and leaders believed that REDCap served an important role for program operations, especially after they implemented strategies to improve the quality of data entered into it. In the second program year, for example, program management began conducting daily quality assurance checks on recently entered data, both to make corrections and to provide feedback on data completeness to staff conducting the home visits. Yale University also reduced the number of free text response options in REDCap to improve the consistency of collected data, and enhanced other fields used for communication between paramedics and VNA nurses (for example, to indicate if the participant has an aggressive pet, or specific instructions on how to enter the home).

After experiencing low uptake of VNA home visits during the first two program years, Yale University implemented multiple strategies to overcome participants' mistrust of VNA nurses. The awardee sought to ameliorate participants' concerns about the role of VNA nurses by encouraging PRIDE staff to clarify the VNA nurse role to participants during enrollment and the paramedic home visit and also providing VNA nurses with PRIDE uniforms.

In addition, Yale University implemented strategies to overcome challenges engaging the target population. First, it incorporated translation services into the PRIDE model, which reportedly built trust with Latino and Russian communities. Second, to encourage participants to maintain engagement throughout the duration of enrollment and schedule both home visits, Yale University established a financial incentive that provided a $10 and $15 gift card to a local grocery store to each participant who completed a paramedic and VNA home visit, respectively. Third, by the start of the third program year, the awardee sought to reduce the number of participants who forgot their appointments by requesting that participants schedule the home visits within a few weeks of enrollment, distributing appointment reminder cards for home visits scheduled during the enrollment phase, and making appointment reminder phone calls on the day before and the day of a home visit. Finally, program staff contacted participants who had enrolled in PRIDE but had not completed one or both visits. The awardee reported success from mailing letters each quarter to participants with outstanding visits, noting that about one-quarter of those letters were returned with a request to schedule an appointment.

> "[The Latino population] was a scared, isolated, yet large group—16 percent of the population in New Haven. We had to win their trust. Now, we have a relationship with that community and it's a great one."
>
> —PRIDE leader

Awardee leaders also described the University's standardized hiring process as burdensome and time-consuming, resulting in delays in onboarding new PRIDE paramedics and dissuading some potential new hires from joining the PRIDE program in the first program year. The hiring challenges, compounded by a growth in enrollment, resulted in the PRIDE program being short staffed during the second program year. However, program leaders reported that the hiring challenges abated in the third program year, as management better understood the University's hiring procedures, PRIDE's reputation grew among paramedics in the New Haven area, and the awardee received direct recommendations for potential new staff from fire department chiefs. According to multiple respondents, the high volume of staff applicants allowed the awardee to hire high quality individuals, described by a program leader as "people [who] are personable, outgoing, who care about the community, and want to make a difference." In the eighth program quarter, the awardee hired additional paramedics and partnered with an additional VNA. In interviews, PRIDE management noted that the improved recruiting processes and additional VNA partnership enabled the awardee to address the backlog in scheduling home visits caused by a sudden growth of enrollment in the second program year.

C. Assessment of perceived program effects on the delivery of care and outcomes

Overall, the majority of PRIDE staff believed that the program effectively identified participants' needs and connected them to the appropriate services and supports. Approximately three-quarters of staff who completed the non-clinician survey (73 percent or 16 people) reported

that PRIDE had a positive impact on participants' access to care or services, half (50 percent or 11 people) felt the program positively impacted participants' achievement of health goals, and nearly two-thirds (64 percent or 14 people) reported a positive impact on quality of life. Results from the participant satisfaction survey conducted independently by Yale University corroborated the non-clinician survey findings. The participant survey, fielded between July 2016 and August 2017, found that nearly all participants would recommend the program to a friend.

In interviews, program leaders and staff stated that the intimacy and long duration of PRIDE home visits were particularly beneficial for discovering unmet participants needs and increasing participant engagement, which reduced risk of falls and improved other health outcomes. Visiting participants' homes revealed participants' health and residential needs that are not typically apparent during a physician office visit. Paramedics commonly initiated small changes in the home to reduce fall risks, such as suggesting the removal of throw rugs or making referrals for installation of grab bars in the shower. Moreover, program staff and leaders believed that the duration of the PRIDE home visits (from 45 minutes to two hours, depending on the participants' level of need and desire to talk) led to stronger participant engagement than typical physician office visits. During these visits, paramedics and VNA nurses took the time to learn about participants' concerns that may pose a barrier to care or services; this information enabled PRIDE to connect participants to appropriate resources, such as Meals on Wheels or transportation services to a PCP appointment.

> "The biggest thing that we're able to do…is actually to get inside the home. Doctors…can talk to people about these issues in their office or in the ED, but talking to them, going into the home, and coming up with the list of 'hey, this is not safe.' That, I think, is making a big difference."
>
> —Program management staff

> "I left my full-time position to do this. I think it was a need in the community. It was just another tool to really have people's needs met. I was very, very excited about it."
>
> —VNA nurse

PRIDE leaders at Yale University also said that an internal evaluation of the program that used outcome data to assess the impact of PRIDE found that the program appeared to reduce ED use within 90 days after receiving PRIDE services. In alignment with this finding, a VNA nurse said she observed that participants are less likely to go to the ED to address non-emergent health needs, like lower back pain, after connecting with the PRIDE program, and are more likely instead to use the new connections that PRIDE staff helped establish with their PCPs, home health agencies, or other resources like Agency on Aging to address their concerns. However, Yale University noted that its internal evaluation also found that PRIDE appeared to have a more substantial impact on enrollees recruited in the ED, compared to participants enrolled via community-based efforts, either as a self-referral or after a 911 call.

D. Implications of implementation findings for the design and interpretation of an impact analysis

This section highlights the implications of our findings for the design and interpretation of an impact analysis.

The change in PRIDE's eligibility criteria after the first program year may have limited the program's effect on desired outcomes (for example, reductions in lift assist calls, mortality, ED visits, hospital admissions, and total expenditures in the target population), and thus reduced the potential to identify improvements in desired outcomes in an impact evaluation. The original recruitment strategy focused on individuals who called 911 for a lift assist after experiencing a fall. The awardee believed these participants were at risk of falling again and would benefit from home assessments and renewed connections with their health care providers. By expanding the eligibility criteria to enroll individuals who were afraid of falling or who providers believed were at risk of falling, the awardee may have included participants with lower risk of injury and, therefore, less potential for improved health outcomes. As previously reported, almost all participants (98 percent) were recruited in the ED or via a self-referral, rather than after a 911 lift assist call, and thus may not have experienced a recent fall, or have ever fallen. PRIDE did not collect information on whether or not enrolled participants had experienced a recent fall or had ever fallen.

Similarly, low enrollment and limited uptake of program services in the first program year may also mute the potential to assess the impact of the PRIDE program. Although enrollment rates increased substantially in the end of the second program year, low enrollment rates in the first program year affected both the size of the sample available for the impact analysis and the ability of an impact analysis to assess the longer-term impact of the program for the majority of participants who were enrolled for a year or less. In addition, almost half of participants did not receive a home visit from a VNA nurse by the end of the third program year and, thus, did not have an opportunity to benefit from the VNA nurse's assessment of the participant's function and mobility or linkage to relevant available resources and services. The no-cost extension period may enable Yale University to complete more VNA nurse home visits.

Another challenge to designing an analysis of the impact of the PRIDE program is the difficulty of identifying an equivalent comparison group for PRIDE participants. Typical data sources, such as Medicare FFS data or hospital electronic health records, are insufficient for capturing the strategies and inclusion criteria used to identify participants enrolled through the ED, after a 911 call, and based on self-referral. In addition, these data sources do not include variables capturing an individual's fear of falling, which is a key eligibility criterion for the PRIDE program and a factor likely associated with impact findings.

Finally, health care outcomes, utilization, and costs may improve in the long term for some PRIDE participants, but potentially not during the three-year cooperative agreement. An evaluation of PRIDE's impact may need to track participants outcomes over a much longer time frame than three years to observe differences in ED visits, hospitalizations, total cost of care, and mortality, because it may take time for participants to fall (or to avoid falling) compared to their counterparts who did not participate in the program. A longer time frame to evaluate health care outcomes may be particularly important for the potentially healthier participants that Yale University recruited through community events or self-referral mechanisms.

III. FINDINGS FROM THE IMPACT EVALUATION

A. Summary of program evaluability

We conducted a detailed reassessment of the evaluability of each of the remaining 38 HCIA R2 awardees—that is, the extent to which we believe we can produce meaningful estimates of program effects on Medicare and Medicaid spending, hospital admissions, unplanned readmissions, ED visits, and other important outcomes specific to each awardee. Our primary assessment of evaluability focused on three evaluability elements presented in Table III.1: (1) whether the projected enrollment at the end of each cooperative agreement is large enough to detect a 20 percent effect on expenditures, which is a large program effect; (2) whether claims can identify the primary expected effects; and (3) whether we can identify a credible comparison group by using claims data.

Yale University received a 12-month no-cost extension and will continue to operate until August 31, 2018, although enrollment of new participants will end February 28, 2018. The projected sample sizes in Table III.1 are based upon actual enrollment as of August 31, 2017 but projected through February 28, 2018 to allow for all participants to receive six months of program exposure, a requirement for inclusion in our evaluation to minimize dilution of program effects due to limited exposure to the program components. Due to processing lags in Medicaid data, we have not confirmed that the 281 Medicaid beneficiaries meet program eligibility for inclusion in any analysis. Further, the projections include all participants in the PRIDE program; however, if sample size is sufficient, our evaluation will be restricted to one arm of the program, ED visits, to allow us to find a comparison group based upon an event captured in Medicare claims data.

Table III.1. Assessment of HCIA R2 awardee evaluability, as of August 31, 2017: Yale University

Evaluability domain	Response
Projected Medicare fee-for-service (FFS) population with 6 months of program exposure as of February 28, 2018	1,186[a]
Projected Medicaid population with 6 months of program exposure as of February 28, 2018	281[a]
Minimum detectible effect (MDE) sample size requirement to detect 10% effect	
Total expenditures	1,437
Likelihood of all-cause hospitalizations	861
MDE sample size requirement to detect 20% effect	
Total expenditures	359
Likelihood of all-cause hospitalizations	215
Participation/selection bias of concern	Limited or no concern
Intervention implemented that differs from baseline period	Fully implemented new intervention relative to baseline

Table III.1 *(continued)*

Evaluability domain	Response
Claims sufficient to identify treatment and comparable comparison group?	Questionable, moderate rate of identification of treatment group with claims with some dilution of treatment effect within an ITT framework
Likelihood of solid comparison group	Serious concern. We may not be able to identify a strong comparison group.
Do claims identify the primary expected effects?	Yes
Core outcomes estimation method	None
Primary reason for no rigorous evaluation	Inability to identify a strong comparison group
Survey data for treatment group that will be analyzed	Staff and beneficiary surveys
Implementation data that will be analyzed	Awardee has provided some participant-level data about the receipt of intervention services (paramedic visit and VNA visit).

[a] The number of enrollees in our impact analysis will be different from those reported in the implementation chapter due to the inclusion criteria used to define the treatment group.

At this point, we are assessing the feasibility of conducting a rigorous impact analysis. Although the number of participants in the ED visit recruitment arm has increased significantly, so has the variation in the strategies that the awardee is using to identify participants. The new recruitment criterion for identifying potential participants includes ED patients who appear to be at risk for falling; and the identification of potential participants is based largely on the perception of ED clinical staff and PRIDE research staff. We will conduct additional exploratory analyses to determine whether we can identify a valid comparison population of ED patients. Absent an impact evaluation, we will report on the rate of receipt of intervention services (for example, paramedic and VNA visits) from participant-level data that the awardee has provided. We will also report on the experiences of awardee staff and participants, based on our surveys.

B. Characteristics of Medicare and Medicaid participants at baseline

This section presents baseline characteristics of Medicare fee-for-service (FFS) beneficiaries in each of the three recruitment methods: ED visit, 911 lift assist, and self-referral. We measured the characteristics during the 12 months before each beneficiary's enrollment into the intervention; enrollment is defined as the beneficiary consenting to participate in the intervention. The characteristics include demographic information, as well as summary statistics on expenditures and the health care utilization of beneficiaries by recruitment method. Future analyses will focus on participants recruited through ED visits only because our ability to identify a comparison group for the 911 lift assist and self-referral recruitment methods is limited. For the purpose of our evaluation, the treatment group consists of individuals enrolled in Medicare FFS, Medicaid, or both who received services for a fall-related event in the ED of Yale New Haven Hospital.[4] Therefore, we focus our description of baseline characteristics on ED visit participants and how beneficiaries from the other two recruitment methods compare with ED

[4] Yale University expanded its efforts to recruit participants from the ED at Yale New Haven Hospital's St. Raphael campus in February 2016.

visit participants. Because there are few Medicaid patients enrolled through the ED, we do not report their baseline characteristics.

Yale University began to enroll participants in the PRIDE program in April 2015. As of May 31, 2016, the awardee had enrolled 569 unique participants in the program.[5] The majority of participants (543 individuals) are Medicare or Medicaid beneficiaries. There are 519 Medicare participants and 282 Medicaid participants (258 individuals participate in both Medicare and Medicaid). The remaining 5 percent (26 individuals) either have other sources of health care coverage or they are uninsured; these participants are not included in our analysis.

In presenting the baseline characteristics, we restricted the treatment group to Medicare beneficiaries who were enrolled in Medicare FFS, both Parts A and B, with Medicare as the primary payer when their eligibility for awardee-provided services began (that is, their enrollment date) and who met all program criteria for a period of 90 days during the baseline year (the 365 days immediately before their enrollment). In addition, they must have been enrolled in the awardee's program on or before May 31, 2016. The calendar period covered by the baseline quarters is based on the enrollment date for each participant and will therefore vary by participant. After we excluded patients who did not meet the above criteria, a total of 329 Medicare FFS beneficiaries were included in the analysis of baseline characteristics for this report. One hundred ninety-three beneficiaries were recruited via the ED visit method, 45 beneficiaries were recruited via the 911 lift assist method, and 95 beneficiaries were recruited via the self-referral method (one beneficiary was enrolled through both a 911 lift assist and an ED visit, one beneficiary was enrolled through an ED visit and self-referral, and two beneficiaries were enrolled through a 911 lift assist and self-referral).

The Medicare FFS beneficiaries recruited by Yale University through the ED are predominantly elderly, female, and white (Table III.2). Ninety percent of them are older than 65, and 44 percent are older than 85. Like Connecticut's Medicare FFS beneficiaries, 80 percent of whom are white, and 10 percent are black,[6] the ED visit beneficiaries are predominantly white (79 percent); 17 percent of them are black. Compared with 16 percent of Medicare beneficiaries nationwide and 13 percent of Medicare beneficiaries in Connecticut,[7] 22 percent of ED visit

[5] The enrollment data presented in Chapter II, Figure II.1, were reported by the awardee to the implementation and monitoring contractor through August 31, 2017. In contrast, the enrollment data in Chapter III are from finder files furnished by the awardee to us. The number of participants in the finder files differ from the number provided to the implementation and monitoring contractor for several reasons. First, the finder file was submitted to us before August 31, 2017, so it does not include all participants. In addition, we restricted finder file enrollment to participants who met additional criteria for our analyses (for example, participants who are in Medicare FFS or in Medicaid and those participants for whom the finder file has valid Medicare or Medicaid identification numbers that enable us to locate the participants' claims data). The impact analyses are further limited to participants for whom we have at least six months of post-enrollment data.

[6] Kaiser Family Foundation, "Distribution of Medicare Beneficiaries by Race/Ethnicity in 2015." Available at http://kff.org/medicare/state-indicator/medicare-beneficiaries-by-raceethnicity/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D. Accessed October 2016.

[7] Kaiser Family Foundation, "Distribution of Medicare Beneficiaries by Eligibility Category in 2013." http://kff.org/medicare/state-indicator/distribution-of-medicare-beneficiaries-by-eligibility-category-

beneficiaries were originally eligible for Medicare because of a disability. Two percent were entitled to Medicare because of end-stage renal disease (ESRD). Compared with 21 percent of Medicare beneficiaries nationwide and 24 percent of Medicare beneficiaries in Connecticut,[8] ED visit beneficiaries who are dual eligibles represent 34 percent of all ED visit beneficiaries, reflecting a high level of social need. ED visit beneficiaries have a mean hierarchical condition categories (HCC) risk score of 2.32 (relative to a national mean risk score of 1.00), reflecting their poorer health status and greater needs for care than the general Medicare FFS population.

The Medicare FFS beneficiaries recruited by Yale University through 911 lift assist are more similar to ED visit beneficiaries than self-referral beneficiaries. Compared with ED visit beneficiaries, 911 lift assist beneficiaries are more likely to be dual eligibles (44 percent compared with 34 percent of ED visit beneficiaries); they also have a higher mean HCC score (2.63). In contrast, self-referral beneficiaries are more likely to be female (81 percent) and white (87 percent), and to have a lower mean HCC score (1.72).

Table III.2. Baseline year demographic characteristics of Medicare FFS beneficiaries enrolled in the treatment group of Yale University's program through May 31, 2016

	ED visit (N=193)		911 Lift Assist (N = 45)		Self-Referral (N=95)	
Characteristics	**Number**	**Percentage**	**Number**	**Percentage**	**Number**	**Percentage**
Age as of enrollment date						
Younger than 65	20	10	6	13	13	14
65 to 74	45	23	7	16	14	15
75 to 84	43	22	15	33	26	27
85 and older	85	44	17	38	42	44
Gender						
Female	127	66	27	60	77	81
Male	66	34	18	40	18	19
Race						
White	153	79	38	84	83	87
Black	32	17	7	16	12	13
American Indian, Alaska Native, Asian/Pacific Island American, or other	4	2	0	0	0	0
Hispanic	4	2	0	0	0	0
Original reason for Medicare eligibility						
Old age and survivor's insurance	148	77	32	71	71	75

2/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D. Accessed October 2016.

[8] Kaiser Family Foundation, "Dual Eligibles as a Percent of Total Medicare Beneficiaries in FY 2011." Available at http://kff.org/medicaid/state-indicator/duals-as-a-of-medicare-beneficiaries/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D. Accessed October 2016.

Table III.2 *(continued)*

Characteristics	ED visit (N=193)		911 Lift Assist (N = 45)		Self-Referral (N=95)	
	Number	Percentage	Number	Percentage	Number	Percentage
Disability insurance benefits	42	22	11	24	23	24
End-stage renal disease (ESRD)[a]	3	2	2	4	1	1
Hospice[b]	0	0	0	0	0	0
Medicare/Medicaid dual status, percentage dual[b]						
HCC score[c]						
Mean		2.32		2.63		1.72
25th percentile		1.03		1.21		0.87
Median		1.77		2.17		1.37
75th percentile		3.02		3.6		2.32

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Note: The baseline year is defined as the 365 days before each beneficiary's enrollment date. The enrollment date is defined as the date on which the participant signs a consent form after being identified through one of three recruitment methods (911 lift assist, ED visit, and self-referral). For future reports, the enrollment date will be the date on which we first have evidence that a beneficiary received treatment in the Yale New Haven Hospital ED for a fall-related event. All beneficiary characteristics were measured during or as of the end of the baseline year.

[a]Includes participants with both a disability and ESRD.

[b]Identified in the last month of each beneficiary's baseline year.

[c]We calculated HCC scores by using the most recently available HCC algorithms developed by CMS. For participants who reside in a long-term care or similar facility, we used the HCC institutional algorithm. For participants enrolled for 12 or fewer months as of the start of their enrollment, we used the new enrollee algorithm. For participants with ESRD, we used the ESRD HCC algorithm and assigned the HCC score based on the participant's age and whether he or she had a transplant, was in an institution, or was a new enrollee. We used the HCC community algorithm for all other enrollees.

FFS = fee-for-service; HCC = hierarchical condition category.

Consistent with the high HCC scores, ED visit beneficiaries had higher expenditures in the year prior to enrollment. In Table III.3, we report baseline utilization and expenditure data for a common set of measures, including the four CMMI core measures. Yale University expects to reduce ED visits and hospital admissions within 30 days of the intervention. We examined the baseline cost of care by calculating average per beneficiary per month (PBPM) Medicare payments in total and by major types of services. Relative to the Connecticut Medicare FFS beneficiaries' average of $917 in 2014,[9] the total average PBPM Medicare payment during the baseline year was $3,004 for ED visit beneficiaries. As expected, the average PBPM payment in the fourth quarter of the baseline year was the highest quarterly PBPM payment because it included the recruitment ED visit; total PBPM payments ranged from $2,281 to $4,560. For ED visit beneficiaries, the average PBPM Medicare payment for acute inpatient care ($1,372) was the largest driver of the total cost of care; this payment is almost 50 percent of the total cost of care. Average PBPM payments for physician services ($480), outpatient services ($410), home health services ($347), and skilled nursing facilities ($334) each contributed approximately 10 to 15 percent of the total cost of care. Quarterly expenditures for acute inpatient services were relatively stable in the first three baseline quarters and increased significantly in the fourth

[9] Unless otherwise noted, national and state data in this paragraph are from the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2016. Available at http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf?sfvrsn=0.

quarter of the baseline year. Quarterly expenditures for physician, outpatient, home health, and skilled nursing facility services were relatively stable over time, although all average payments were higher in the fourth quarter of the baseline year.

Compared with ED visit beneficiaries, the total average PBPM Medicare payment during the baseline year was a little higher for 911 lift assist beneficiaries ($3,442) and much lower for self-referral beneficiaries ($1,352). The quarterly average PBPM estimates for 911 lift assist beneficiaries increased over time, whereas these estimates were stable for self-referral beneficiaries. Compared with ED visit beneficiaries, for whom acute inpatient services were almost 50 percent of total cost of care in the baseline year, payments were more equally distributed across services for 911 lift assist and self-referral beneficiaries.

The baseline rate of acute care hospitalizations for ED visit beneficiaries was 1,203 per 1,000 Medicare FFS beneficiaries in the baseline year—higher than the Connecticut average in 2014 of 291 per 1,000 Medicare FFS beneficiaries.[10] Fifty-six percent of ED visit beneficiaries had at least one hospitalization during the year before enrollment. The rate of outpatient ED visits for ED visit beneficiaries (1,547 per 1,000 Medicare FFS beneficiaries in the baseline year, or 65 percent of beneficiaries) was more than three times higher than the 2013 national rate of 454 per 1,000 Medicare FFS beneficiaries.[11] The higher rate of outpatient ED visits among beneficiaries suggests that there is an opportunity to reduce the rate of outpatient ED visits through PRIDE program services. The baseline rate of ambulatory observation stays (607 per 1,000 beneficiaries per year) was more than 10 times greater than the 2014 national average of 58 per 1,000 beneficiaries.[12] The percentage of discharges with a 30-day readmission among ED visit beneficiaries (20 percent per discharge) in the baseline year was slightly higher than the Connecticut average percent in 2014 for Medicare beneficiaries (18 percent per discharge); 11 percent of all ED visit beneficiaries had a readmission during the baseline year. At baseline, the rate of primary care visits in any setting was 10,487 per 1,000 Medicare FFS beneficiaries per year for ED visit beneficiaries. This rate falls by 17 percent to 8,694 per 1,000 Medicare FFS beneficiaries if we restrict the settings to the ambulatory setting. The baseline rate of specialty care service use in any setting was 20,808 per 1,000 Medicare FFS beneficiaries per year; this rate falls by 41 percent to 12,173 per 1,000 Medicare FFS beneficiaries if we look at ambulatory settings only. Similar to the trend observed in expenditures for ED visit beneficiaries, quarterly rates for most utilization measures were markedly higher in the fourth quarter of the baseline year than in other quarters. Primary care and specialist visits in ambulatory settings and the percentage of readmissions among

[10] Unless otherwise noted, national and state data in this paragraph are from the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2016. Available at http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf?sfvrsn=0. Accessed August 2016

[11] National outpatient ED rate calculated from the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2016. Available at http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf?sfvrsn=0. Accessed August 2016.

[12] See the Medicare Payment Advisory Commission, "A Data Book: Health Care Spending and the Medicare Program," June 2016. Available at http://www.medpac.gov/docs/default-source/data-book/june-2016-data-book-health-care-spending-and-the-medicare-program.pdf?sfvrsn=0. Accessed August 2016.

all ED visit beneficiaries were relatively stable across the four baseline quarters, whereas the percentage of 30-day readmissions trended downwards across the four quarters.

Compared with ED visit beneficiaries, most utilization of hospital-related services (including acute hospital admissions, outpatient ED visits, and observation stays) was approximately one-third lower for 911 lift assist beneficiaries and approximately two-thirds lower for self-referral beneficiaries. The percentage of discharges with a 30-day readmission was lower for 911 lift assist beneficiaries (10 percent) and self-referral beneficiaries (14 percent) than it was for ED visit beneficiaries (20 percent). However, a greater percentage of self-referral beneficiaries had a 30-day readmission than did 911 lift assist beneficiaries. The latter had rates of primary care and specialty care visits that were similar to those of ED visit beneficiaries. Self-referral beneficiaries had lower rates of primary care visits in any setting, primary care visits in ambulatory settings, and specialty visits in any settings; however, rates of specialty visits in ambulatory settings were similar for self-referral and ED visit beneficiaries. No distinct pattern in quarterly rates was observed for 911 lift assist or self-referral beneficiaries.

Overall, beneficiaries in the PRIDE program who enrolled through an ED visit had higher expenditures and utilization relative to the national and Connecticut averages for all Medicare FFS beneficiaries, suggesting that there is the potential to improve care for all participating beneficiaries. However, the baseline characteristics, health care expenditures, and utilization rates suggest that the three recruitment arms are enrolling different Medicare FFS populations. The populations recruited through the ED visit and 911 lift assist were more similar to each other in the baseline year, whereas self-referral beneficiaries were healthier and used fewer health care resources. Total expenditures for ED visit beneficiaries were similar to total expenditures for 911 lift assist beneficiaries, but ED visit beneficiaries used more hospital-related services than did 911 lift assist beneficiaries. Self-referral beneficiaries had lower expenditures and utilization rates in all measures except specialist visits in ambulatory settings.

Table III.3. Baseline year expenditures and health care utilization for Medicare FFS beneficiaries enrolled in the treatment group of Yale University's program through May 31, 2016, by quarter and recruitment arm

	Expenditures and utilization for each quarter in the 12 months before and including the enrollment date														
	ED Visit					911 Lift Assist					Self-Referral				
	BL	Q1	Q2	Q3	Q4	BL	Q1	Q2	Q3	Q4	BL	Q1	Q2	Q3	Q4
Total number of enrollees	193	183	186	192	193	45	45	45	45	45	95	93	93	95	95
Average Medicare expenditures PBPM[a]															
Total	3,004 (274)	2,614 (405)	2,512 (382)	2,281 (261)	4,560 (448)	3,442 (589)	1,930 (460)	3,391 (792)	4,140 (1,022)	4,306 (980)	1,352 (219)	1,723 (424)	1,168 (231)	1,351 (391)	1,169 (229)
Acute inpatient	1,372 (181)	1,064 (271)	1,034 (268)	856 (155)	2,498 (328)	870 (223)	308 (157)	866 (293)	1,084 (416)	1,223 (523)	371 (101)	459 (165)	182 (84)	574 (290)	267 (110)
Inpatient other[b]	13 (7)	35 (26)	0 (0)	9 (9)	10 (10)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Outpatient[c]	410 (53)	386 (69)	380 (82)	347 (56)	526 (61)	864 (388)	399 (104)	975 (498)	1,134 (629)	948 (401)	188 (44)	155 (48)	235 (98)	134 (37)	228 (62)
Physician services	480 (29)	447 (44)	391 (34)	415 (36)	659 (51)	570 (110)	658 (223)	530 (126)	540 (106)	551 (116)	345 (36)	391 (48)	363 (43)	341 (45)	285 (38)
Home health	347 (41)	319 (59)	331 (60)	352 (52)	385 (59)	566 (79)	394 (123)	564 (145)	463 (117)	843 (179)	198 (40)	185 (53)	222 (73)	135 (45)	247 (64)
Skilled nursing facility	334 (51)	342 (98)	306 (88)	235 (87)	453 (109)	489 (155)	133 (117)	393 (269)	830 (399)	600 (289)	220 (70)	522 (207)	125 (72)	117 (74)	121 (60)
Hospice	10 (7)	0 (0)	28 (27)	12 (11)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	14 (14)	0 (0)	26 (26)	32 (32)	0 (0)
Durable medical equipment	38 (14)	23 (5)	43 (17)	55 (28)	29 (9)	82 (28)	38 (10)	61 (19)	89 (35)	141 (103)	16 (4)	10 (3)	15 (4)	18 (5)	20 (6)
Health care utilization rates (annualized per 1,000)															
Acute hospital admissions[d]	1,203 (123)	796 (174)	1,042 (184)	891 (140)	2,051 (247)	822 (169)	356 (172)	1,156 (351)	800 (241)	978 (340)	373 (82)	516 (164)	215 (94)	464 (168)	295 (108)
Outpatient ED visits	1,547 (239)	1,215 (409)	1,042 (250)	954 (242)	2,930 (288)	956 (171)	800 (273)	889 (309)	711 (263)	1,422 (385)	607 (111)	603 (162)	774 (239)	590 (158)	463 (145)
Observation stays	607 (83)	530 (137)	261 (84)	488 (112)	1,130 (271)	422 (140)	444 (190)	267 (197)	356 (214)	622 (335)	181 (60)	258 (119)	215 (94)	0 (0)	253 (100)

Table III.3 *(continued)*

	Expenditures and utilization for each quarter in the 12 months before and including the enrollment date														
	ED Visit					911 Lift Assist					Self-Referral				
	BL	Q1	Q2	Q3	Q4	BL	Q1	Q2	Q3	Q4	BL	Q1	Q2	Q3	Q4
Primary care visits in any setting	10,487 (799)	11,160 (1,246)	8,924 (1,028)	9,735 (927)	12,054 (1,181)	11,289 (1,976)	9,244 (1,948)	12,889 (3,866)	9,689 (1,980)	13,333 (3,259)	6,461 (630)	8,779 (1,623)	6,108 (694)	5,652 (740)	5,347 (551)
Primary care visits in ambulatory settings	8,694 (691)	9,105 (984)	7,947 (952)	8,059 (795)	9,605 (969)	9,467 (1,828)	7,911 (1,849)	11,200 (3,697)	8,622 (1,932)	10,133 (2,301)	5,919 (565)	8,005 (1,583)	5,720 (651)	5,019 (578)	4,968 (510)
Specialist visits in any setting	20,808 (1,497)	20,022 (2,155)	19,302 (1,874)	17,052 (1,577)	26,640 (2,400)	20,778 (2,883)	20,622 (3,773)	20,000 (3,188)	22,044 (3,481)	20,444 (4,037)	14,786 (1,667)	17,042 (2,077)	16,516 (2,406)	14,045 (2,199)	11,453 (1,462)
Specialist visits in ambulatory settings	12,173 (1,058)	11,757 (1,174)	13,288 (1,537)	11,749 (1,221)	11,845 (1,175)	14,556 (1,838)	16,444 (2,733)	14,756 (2,092)	14,844 (1,976)	12,178 (1,793)	12,508 (1,549)	13,513 (1,631)	14,581 (2,358)	11,726 (2,058)	10,105 (1,410)
Measures of any health care utilization															
Percentage with a hospital admission[d]	56 (4)	13 (2)	19 (3)	19 (3)	39 (4)	44 (7)	9 (4)	22 (6)	20 (6)	20 (6)	22 (4)	11 (3)	5 (2)	8 (3)	7 (3)
Percentage with an outpatient ED visit	65 (3)	17 (3)	15 (3)	15 (3)	54 (4)	56 (7)	18 (6)	18 (6)	16 (5)	27 (7)	36 (5)	14 (4)	13 (3)	14 (4)	11 (3)
Percentage with an observation stay	38 (4)	10 (2)	5 (2)	11 (2)	23 (3)	27 (7)	11 (5)	4 (3)	7 (4)	9 (4)	13 (3)	5 (2)	5 (2)	0 (0)	6 (3)
Percentage with a 30-day readmission among all discharges	20 (3)	27 (7)	22 (6)	18 (6)	13 (5)	10 (5)	20 (20)	0 (0)	8 (8)	20 (13)	14 (6)	15 (10)	0 (0)	17 (17)	17 (11)
Percentage of participants with a readmission among all participants	11 (2)	4 (1)	4 (1)	3 (1)	3 (1)	7 (4)	2 (2)	0 (0)	2 (2)	2 (2)	4 (2)	2 (2)	0 (0)	1 (1)	1 (1)

Source: Mathematica analysis of information from awardee's finder file and Medicare claims and enrollment data as of May 31, 2016

Notes: The baseline year is the 366 days before and including each participant's enrollment date. Each baseline quarter is defined as a 91- or 92-day period starting from each beneficiary's enrollment date. The fourth baseline quarter is the 92 days before and including each beneficiary's enrollment date, the third baseline quarter is the 91 days before the fourth baseline quarter, the second baseline quarter is the 91 days before the third baseline quarter, and the first baseline quarter is the 92 days before the second baseline quarter, thereby summing to 366 days. We weight every outcome during the baseline year and each baseline quarter according to the number of days each beneficiary is enrolled in FFS Medicare during the baseline year and in each baseline quarter.
Standard errors are shown in parentheses.

[a]Total Medicare expenditures for the baseline year or a given quarter were calculated from all claims for each participant with at least one eligible day during that year or quarter.

[b]Inpatient "other" expenditures include the other types of inpatient stays, such as rehabilitation admissions, long-term care hospital services, or psychiatric hospital services.

[c]Includes visits to an ED and a hospital outpatient department, as well as outpatient surgeries.

[d]The hospitalization and readmission measures include acute care hospital admissions and exclude all other types of rehabilitation admissions, long-term care hospital services, or psychiatric hospital services. The number of acute care hospital admissions was calculated from claims data for each participant with at least one eligible day during that quarter.

BL = baseline year average; ED = emergency department; FFS = fee-for-service; PBPM = per beneficiary per month.

This page has been left blank for double-sided copying.

IV. FINDINGS FROM THE EVALUATION OF THE PAYMENT MODEL

Yale University designed a payment model in which the regional EMS medical director receives prospective payments for a geographical region. The medical director serves as an intermediary and reimburses paramedics, visiting nurses, and scheduled transportation providers for providing specific services, including an initial in-home assessment by the responding paramedic, a two-hour initial home visit by a visiting nurse, and scheduled round-trip transportation to the participant's PCP.

A. Data and methods for evaluating the payment model

The findings in this chapter are based on our synthesis of data from the following sources: (1) the awardee's final payment model report submitted in June 2017 to the implementation and monitoring contractor and (2) narratives and payment model reports the awardee submitted to the implementation and monitoring contractor. We did not interview the awardee about its payment model because it did not advance from the development stage to active negotiations with one or more payers.

B. Description of the payment model

Under Yale University's payment model, the regional EMS medical director receives a prospective, population-based payment for all beneficiaries in a geographic region to proactively address fall risk for individuals who have already fallen or are at risk of falling in their homes. The payment provides the medical director with funds to reimburse paramedics, visiting nurses, and scheduled transportation providers for conducting home assessments and linking participants to their PCPs. The awardee emphasized the importance of an appropriate per beneficiary per month (PBPM) payment that considers anticipated costs, which could be calculated based on an actuarial analysis and using data collected during the three-year cooperative agreement. The awardee also noted that EMS agencies would find that implementing the program reduces the number of lift assist calls, for which EMS agencies do not receive reimbursements.

C. Status of the payment model

Yale University reported engaging Medicare in discussions about the PRIDE program, given that the program's target population is elderly patients with mobility impairment. In addition, the awardee noted having little success engaging private payers early on during the cooperative agreement period. At the end of the third program year, the awardee began drafting a proposal to submit to the Physician-Focused Payment Model Technical Advisory Committee (PTAC) that described how PRIDE services could be provided through an alternative payment model. The awardee planned to use the no-cost extension period to complete that proposal, including demonstrating the quality and cost-savings impact of the PRIDE program and developing a partnership with Yale New Haven Hospital. Yale University did not plan to implement the payment model during the no-cost extension period.

D. Factors associated with the development of the payment model

Yale University reported data availability as the primary impediment in developing its payment model. Although the awardee calculated a preliminary PBPM amount, it noted that a revised calculation was necessary to ensure the estimated PBPM provided the EMS medical

director with sufficient funds to cover program services. The PBPM calculation required an estimate of the expected number of people who would use PRIDE services in a given region and expected cost savings that would result from implementing the program in that given region. Given the growth in enrollment throughout the cooperative agreement, the awardee planned to complete analyses during the no-cost extension period using the most recent participant data and insight from a health economist. Also during the no-cost extension period, the awardee expected to incorporate feedback from paramedics, program staff, enrolled participants, and VNA nurses to ensure the payment model included an appropriate balance of financial incentives (such as salary) and professional incentives (such as EMS personnel's interest in avoiding repeated low-acuity 911 calls).

V. FINDINGS FROM THE ASSESSMENT OF PROGRAM SUSTAINABILITY, SCALABILITY, AND REPLICABILITY

Yale University plans to use the no-cost extension period to develop a sustainability plan and transition the program from an academic setting to the public/private sector. Specifically, Yale University will continue enrolling and providing program services, engaging payers after collecting and analyzing data to inform payment model development, and considering regulatory options to address the paramedics' scope of practice requirements. Yale University did not report plans to scale or replicate its program.

A. Data and methods for assessing program sustainability, scalability, and replicability (SSR)

This chapter of the narrative summarizes the awardee's plans to continue and expand its program after the cooperative agreement through three separate activities: sustainability, scalability, and replicability, as defined below:

- Sustainability refers to the continuation of the program components, in whole or in part, in the same provider organizations and target populations.
- Scalability refers to the expansion of the program to other populations, providers, sites of care, geographical areas, or payers beyond the original plans and enrollment targets. Scaling does not include changes in eligibility criteria, providers, or catchment areas as a strategy to meet enrollment targets for the cooperative agreement.
- Replicability refers to the adoption of the program, in whole or in part, by an organization other than the awardee for individuals who are not participating in the HCIA R2 program.

We used several data sources to assess SSR. The main source of information on the awardee's SSR progress (Section C) and the factors affecting progress (Section D) was the implementation interviews conducted with program leaders, staff, and partners during the summer of 2017. Those interviews included questions about progress in sustaining, scaling and replicating the programs, and provided respondents our definition of these terms. In addition, we used awardee-submitted reports from the last program quarter for updated information about the awardee's SSR activities. To provide context and an updated timeline for the awardee's SSR planning, we reviewed information from CMS on which awardees received a no-cost extension and the period and purpose of these extensions. Before discussing the awardee's progress, we summarize the awardee's baseline SSR plans from the second year of the program (Section B).

B. Baseline SSR plan

As of the second program year, Yale University reported plans to convene a planning committee to develop a sustainability plan for PRIDE. During the first two program years, the awardee reported focusing on recruiting a sufficient number of participants to support analyses of the program that could be used to engage potential payers and develop the details of the payment model.

C. Implementing the SSR plan: progress and changes

Sustainability. As of the third program year, Yale University's sustainability planning committee had started discussing next steps for sustainment planning. The awardee anticipated that the no-cost extension period would provide it with more program data to analyze and use to engage payers, specifically CMS. At the end of the third program year, awardee leaders reported being in the process of contracting with a health economist and hiring two part-time statisticians to help analyze program data, although the awardee did not report what specific analyses they planned to conduct. The awardee received a no-cost extension through August 31, 2018, and will continue to enroll participants through February 2018 and provide program services through May 2018.

Scalability. Yale University did not report plans to scale its program.

Replicability. Yale University did not report any plans to replicate its program.

D. Factors associated with progress toward implementing the SSR plan

The awardee experienced two main challenges to sustainment planning: (1) low enrollment of participants in the first two program years and (2) a state law that restricts community paramedicine. First, Yale University reported that low enrollment during the first two program years meant that it had access to data for a relatively small number of participants who enrolled early in the cooperative agreement period; these did not provide enough sample size to calculate rates and assess the program's impact on health care utilization and costs. Awardee leaders stated that the delay in accessing a sufficiently large sample size limited the awardee's ability to demonstrate the value of PRIDE to prospective payers during the three-year cooperative agreement.

Second, the awardee explained that Connecticut regulations that defined paramedics' scope of practice requirements, including a prohibition on community paramedicine, had impacted the awardee's ability to transition the program from the academic setting to the public/private sector. Operating as a research study during the cooperative agreement, the awardee hired paramedics as Yale University employees, which enabled them to complete the in-home assessments despite the regulatory restriction. In order to sustain the program after the cooperative agreement, the awardee must develop an alternative approach for home visits by paramedics. The awardee reported plans to closely follow ongoing legislative efforts to expand paramedics' scope of practice after the cooperative agreement period ended.

VI. NEXT STEPS FOR THE EVALUATION

A. Implementation evaluation

Although the cooperative agreement has ended, the awardee received a no-cost extension that allows it to continue to enroll new beneficiaries. The implementation team will follow up with the awardee in 2018 to assess whether there are additional data that provide evidence of the program's implementation effectiveness. In addition, should CMS request it, we will follow up with the awardee to discuss the status of the payment model or sustainability, scalability, or replicability plans.

We will continue to review reports, including the final report, submitted by the awardee to the implementation and monitoring contractor to determine whether there is new information that will give us additional insights into the awardee's implementation experience and effectiveness. In addition, we will review the results of the survey administered to participants in 2017 to assess their experience with the program and their perceptions of its effect on the care they received. Finally, the implementation evaluation team will work with the impact team to provide insight into the awardee's implementation experiences that will facilitate the interpretation of the impact analyses. We will include summaries of our findings from these efforts in the final report.

B. Impact evaluation

As noted above, we do not believe that an impact evaluation is feasible. Instead, we plan to develop a descriptive analysis of the program and its outcomes. At this point, we have not spelled out the analysis in detail. We expect either to compare program outcomes to national or regional benchmarks or to contrast outcomes with those of a comparison group that broadly resembles program participants but does not resemble them closely enough to support an impact evaluation. During the early months of 2018, we will specify an approach and supporting data, describing the analysis and the information it is likely to provide.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD