CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 5 Evaluation & Monitoring Annual Report

Prepared for:

CMS

Prepared by:

The Lewin Group

With our partners:

Abt Associates, GDIT, and Telligen

Final

October 2018

CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 5 Evaluation & Monitoring Annual Report

The Lewin Group

With our partners Abt Associates, GDIT, and Telligen

Authors:

Laura Dummit, Grecia Marrufo, Jaclyn Marshall, Tristan Ackerman, Sarah Bergman, Aylin Bradley, Rebecca Cherry, Andrea Chung, Inna Cintina, Daniel Cooper, Syvart Dennen, Ayah Fannoun, Gina Gerding, Dan Gregory, Megan Hyland, Ashley Johnson, Susan Joy, Jordan Kahn, Jon Kelly, Alex Lampert, Karla Lopez de Nava, Brandon Maughan, Court Melin, Sebastian Negrusa, William Sheahan, Julie Somers, Aditya Subramanian, Katarina Swanson, Ellen Tan, Kyi-Sin Than, Peter Weidner, David Zhang, Dean Farley, Ian Breunig, Andrea Hassol, Qian Gu, Sean McClellan, Matt Trombley, Christine LaRocca, Becky Blystone, Colleen Kummet

Lewin's address:

3130 Fairview Park Dr, Suite 500, Falls Church, VA 22042

Federal Project Officer:

Daver Kahvecioglu
Division of Data, Research, and Analytic Methods (DRAM)
Research and Rapid Cycle Evaluation Group (RREG),
Center for Medicare and Medicaid Innovation (CMMI),
Centers for Medicare and Medicaid Services (CMS)

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2011-00001I Task Order HHSM-500-T0007.

The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. The Lewin Group assumes responsibility for the accuracy and completeness of the information contained in this report.

Table of Contents

List of Appendices

Executive Summary

A. Introduction

The Centers for Medicare & Medicaid Services (CMS) implemented the Bundled Payments for Care Improvement (BPCI) initiative under the authority of the Center for Medicare & Medicaid Innovation (CMMI) to test whether linking provider payments for an episode of care could reduce Medicare payments while maintaining or improving the quality of care. The BPCI initiative is comprised of four Models; this evaluation contract covers Models 2, 3, and 4.[1] This summative evaluation incorporates all analyses conducted during the five year evaluation contract and describes the experience under BPCI for over three years of the initiative, from the fourth quarter (Q4) of 2013 through Q4 2016.

BPCI is a voluntary initiative that allows participants to choose among several key design options. This design implicitly recognizes the variability across health care markets, providers, and episodes of care. The resulting diversity in responses and impacts provides CMMI with information on the approaches that show the most promise in achieving payment reductions while maintaining or improving quality. The BPCI initiative rewards participants financially for reducing Medicare payments for an episode of care relative to a target price. Awardees' agreements with CMS specified their Model choice as well as choices among 48 clinical episodes, other episode characteristics, and multiple options for program rule waivers and financial arrangements with other parties. The clinical episodes are defined by the Medicare Severity-Diagnosis Related Group (MS-DRG) of the anchor or qualifying hospitalization. Providers and other organizations that volunteered to participate could enter into the risk-bearing phase of the initiative during a 2-year period through September 2015 and enter additional clinical episodes into the risk-bearing phase through December 2015. Providers can stop participating in a given clinical episode or terminate their participation in the initiative at any time.

Findings in this report are based on analyses of Medicare claims and enrollment data, post-acute care (PAC) provider patient assessments, Awardee-submitted data, beneficiary surveys, participant interviews, and participant site visits. This annual report updates analyses from previous reports and includes new analyses. It is the first to include results of episodes initiated by physician group practices (PGPs) under Model 2. We also estimate the impact of BPCI on quality of care and satisfaction among particular subpopulations of beneficiaries with characteristics that could make them especially vulnerable to changes in care, such as those eligible for both Medicare and Medicaid. We expand our examination of the factors that contribute to whether a participant can reduce episode payments below its target price under BPCI to additional Model 2 hospital and PGP initiated clinical episodes. We refined our methodology to estimate the impact of BPCI on net savings to Medicare. Finally, we assess whether BPCI caused an increase in volume of major joint replacement of the lower extremity (MJRLE) episodes, the highest volume clinical episode in BPCI. Overall, our results are consistent with previous reports that demonstrated that BPCI participants have responded to the initiative's incentives by reducing Medicare payments.

[1] Model 1 began earlier than Models 2, 3, and 4 and was evaluated separately.

B. Results

The average Model 2 episode initiator (EI) participated in eight clinical episodes, and the most commonly selected clinical episode was MJRLE. BPCI Model 2 accounted for nearly 90% of the approximately 796,000 episodes initiated during the first 13 quarters of the initiative. Because providers were allowed to join BPCI over an extended period and stop participating at any time, these data represent an average of seven quarters of participation for hospital EIs and six quarters for PGP EIs. As of December 2016, approximately 20% of hospital EIs and 26% of PGP EIs withdrew completely from the initiative.[2]

Though many more EIs participated in Model 3 than Model 2, episode volume was lower than in Model 2. Skilled nursing facility (SNF) EIs were most likely to participate in MJRLE, where they initiated over 9,600 episodes during the first 13 quarters of the initiative. Congestive heart failure (CHF) had the greatest enrollment of home health agency (HHA) EIs and the largest patient volume, exceeding 4,800 episodes during the same period. The data in this report represent an average of five quarters of participation for SNF EIs and six quarters for HHA EIs. Model 3 participants withdrew from the initiative at similar rates as their Model 2 counterparts, with 27% of SNFs and 30% of HHAs terminating participation in BPCI as of December 2016.[3]

Participation in Model 4 continued to wane in the third year of the initiative. Only five hospitals participated in Model 4 in 2017 and another three Model 4 hospitals transitioned to Model 2 rather than withdraw entirely from the initiative. At the peak of enrollment, 23 episode-initiating hospitals participated in Model 4. A total of 13,551 episodes, primarily for MJRLE, were initiated under the Model through December 2016. Limited sample size made it difficult to draw conclusions about the impact of Model 4 on key outcomes; thus we only present a summary of the previously published results on the impact of Model 4 BPCI in this report. (See the CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 3 Evaluation & Monitoring Annual Report for details about Model 4.[4])

Under the BPCI initiative, Medicare payments declined for most clinical episodes and over half of the relative payment reductions were statistically significant. The declines were primarily due to relative reductions in the use of PAC. The Medicare payment reductions occurred under Model 2 and 3 and across participant types as well as a range of surgical, acute, and chronic clinical episodes. Quality of care, measured as emergency department visits, mortality, and readmissions, was not affected in the vast majority of clinical episodes. Changes in functional status did not differ between beneficiaries in BPCI episodes and comparison beneficiaries, based on survey results, although fewer BPCI beneficiary respondents reported the highest level of satisfaction with their care.

[2] For Model 2, we had sufficient sample size to evaluate 32 hospital-initiated clinical episodes and 21 PGP-initiated clinical episodes with Medicare claims and 25 hospital -initiated and 18 PGP-initiated clinical episodes with the beneficiary survey.

[3] For Model 3, we had sufficient sample size to analyze 11 clinical episodes initiated by SNFs and three clinical episodes initiated by HHAs with claims and three SNF clinical episodes and two HHA clinical episodes with the beneficiary survey.

[4] Available for download at https://innovation.cms.gov/initiatives/bundled-payments/

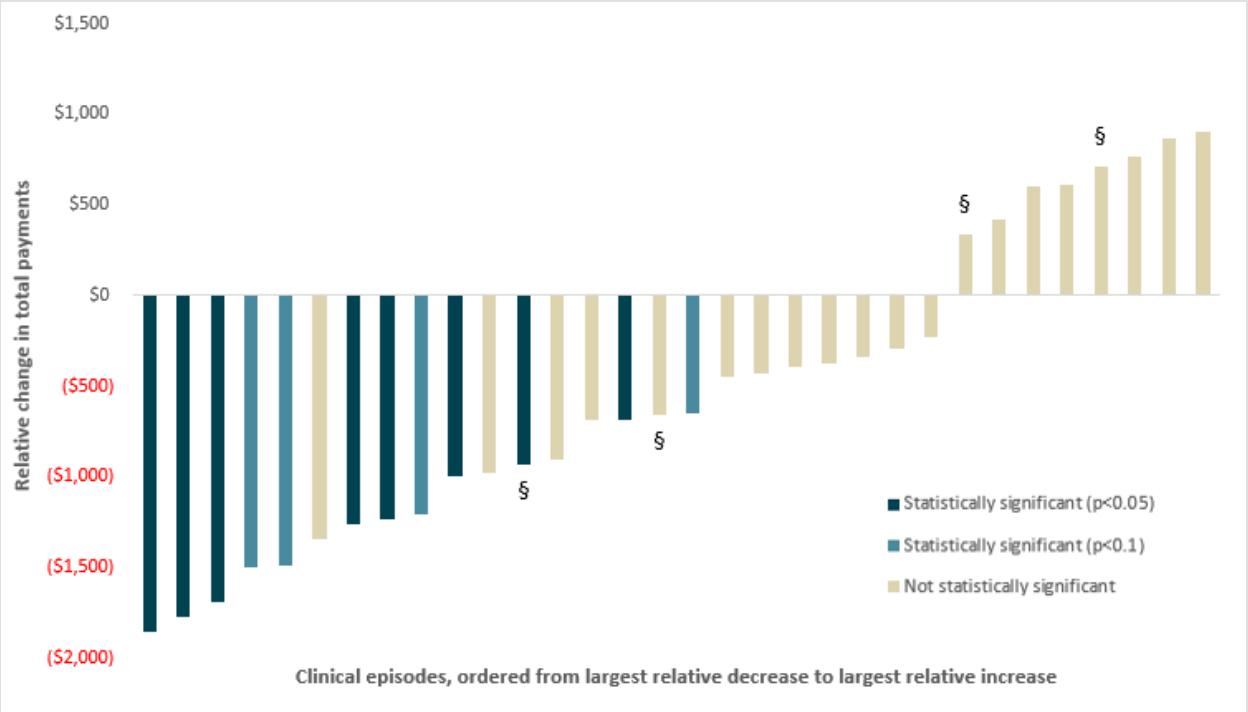
Providers that chose to participate in the voluntary initiative differed from those that did not participate; they tended to have more resources and higher baseline episode payments than non-participants, which may have made it easier to reduce episode spending. BPCI participants did not change their market share of discharges or reduce the number of PACs to which patients were discharged under the initiative. Participants did increase the share of patients discharged to a SNF with a high star rating.

While BPCI was associated with a decline in episode payments, after considering the reconciliation payments made to participants, BPCI did not result in savings to the Medicare program.

- **Medicare payments declined for three-quarters of the clinical episode combinations evaluated with little change in quality of care.**

Across the 67 Model- participant- and clinical episode-combinations analyzed in this report, payments declined for 50 and the change was statistically significant for 27. This is particularly notable because of the consistent pattern of declines across both Models, different participant types, and a variety of clinical episode types (surgical procedures and acute and chronic medical conditions). The total standardized allowed payment amount for the inpatient stay plus 90 days post-discharge had a statistically significant decline for 12 of the 32 hospital-initiated clinical episodes analyzed and for seven of the 21 PGP-initiated clinical episodes ($p<0.10$, Exhibits ES-1 and ES-2). In seven of the 11 SNF Model 3 clinical episodes analyzed, we observed statistically significant reductions in total episode payments and there was a statistically significant decline in payments for one of the three HHA clinical episodes ($p<0.10$, Exhibit ES-3). At the same time, there were few statistically significant changes in quality outcomes; often the statistically significant quality results were not robust, based on various sensitivity tests.

Exhibit ES-1: Impact of BPCI on Total Standardized Allowed Payment Amount for the Inpatient Stay Plus 90-day Post-discharge Period, by Clinical Episode, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016

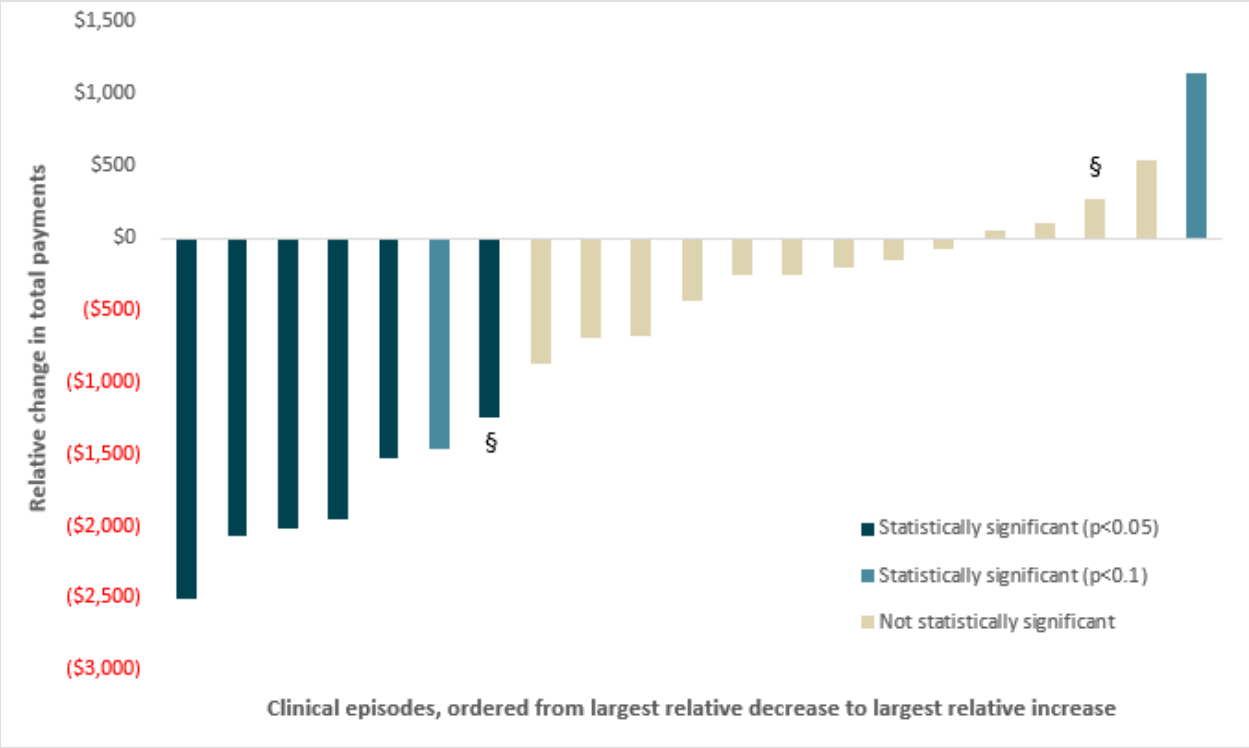


Note: The estimates in this exhibit are the result of a difference-in-differences (DiD) model. These amounts include Medicare program payments plus beneficiary coinsurance and copayments and are adjusted to ensure that any differences across time and providers reflect real differences in resource use rather than Medicare payment policies (e.g., teaching payments or differential payment updates).

§ Data from the baseline period showed BPCI and matched comparison providers were not on parallel trends for this outcome, which is required for an unbiased estimate.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit ES-2: Impact of BPCI on Total Standardized Allowed Payment Amount for the Inpatient Stay Plus 90-day Post-discharge Period, by Clinical Episode, Model 2 PGP, Baseline to Intervention, Q4 2013 – Q4 2016

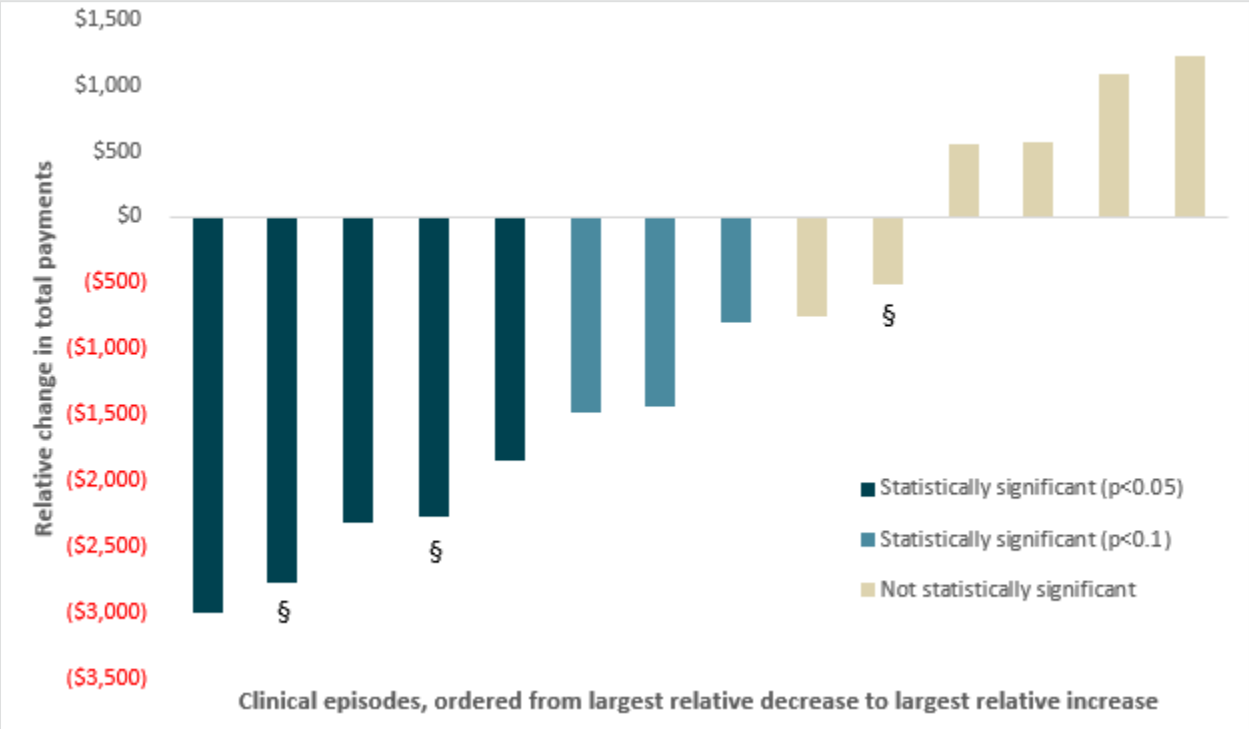


Note: The estimates in this exhibit are the result of a difference-in-differences (DiD) model. These amounts include Medicare program payments plus beneficiary coinsurance and copayments and are adjusted to ensure that any differences across time and providers reflect real differences in resource use rather than Medicare payment policies (e.g., teaching payments or differential payment updates). PGP = physician group practice.

§ Data from the baseline period showed BPCI and matched comparison providers were not on parallel trends for this outcome, which is required for an unbiased estimate.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit ES-3: Impact of BPCI on Total Standardized Allowed Payments Included in 90-Day Bundle Definition, by Clinical Episode, Model 3, Baseline to Intervention, Q4 2013 – Q4 2016



Note: The estimates in this exhibit are the result of a difference-in-differences (DiD) model. These amounts include Medicare program payments plus beneficiary coinsurance and copayments and are adjusted to ensure that any differences across time and providers reflect real differences in resource use rather than Medicare payment policies (e.g., teaching payments or differential payment updates). Data reflect 90-day episodes.

§ Data from the baseline period showed BPCI and matched comparison providers were not on parallel trends for this outcome, which is required for an unbiased estimate.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

- **Lower use of institutional PAC contributed to reduced Medicare payments under Model 2; there were also reductions in HHA payments for PGP-initiated episodes.**

Among the hospital-initiated clinical episodes, reductions in total standardized allowed payment amounts were associated with decreases in institutional PAC payments, particularly SNF payments, and increases in HHA payments. There was a statistically significant decline in SNF payments during the 90-day post-discharge period for 16 of the 32 clinical episodes and an increase in HHA payments for 12 ($p<0.10$).

A lower proportion of the patients who received PAC were discharged to an institutional PAC setting (SNF or inpatient rehab facility (IRF)) for 26 of the 32 hospital-initiated clinical episodes, and this reduction was statistically significant for 13. The number of SNF days for beneficiaries who used SNF care went down for 30 clinical episodes, and the decrease was statistically significant for 17 ($p<0.10$).

PGP-initiated episodes also had reduced payments for institutional PAC. There was a statistically significant decline in SNF payments for five PGP clinical episodes ($p<0.10$). Unlike the hospital-initiated clinical episodes, which often had higher HHA payments, half of the PGP-initiated

clinical episodes had lower HHA payments. The relative decline in HHA payments was statistically significant for four clinical episodes ($p<0.10$).

- **For Model 3 SNF clinical episodes, declines in SNF payments contributed to total payment reductions; payment results were mixed for HHA clinical episodes.**

The statistically significant total episode payment reductions for Model 3 SNF clinical episodes were associated with declines in SNF payments. HHA payments increased for the majority of these clinical episodes, and the increase was significant for four ($p<0.10$). The higher HHA payments, however, did not offset the reduced SNF payments. Higher readmission payments, however, offset the reduction in SNF payments for two clinical episodes.

Among HHA-initiated episodes, there were few statistically significant changes in payments.

- **Claim-based measures generally indicated that quality of care did not change under BPCI Model 2.**

There were few statistically significant changes in mortality, emergency department visits, or unplanned readmissions under Model 2 for either hospital- or PGP-initiated episodes. The point estimates of the differential change in these quality measures tended to be small and in both directions. For the few exceptions to this general finding, additional tests suggested that the results could be biased because BPCI and matched comparison providers did not have parallel trends in the baseline for the measure in question or because of the random selection of episodes included in the matched comparison sample.

- **Self-reported changes in functional status did not differ between beneficiaries with BPCI episodes and comparison beneficiaries; slightly less favorable views of care experiences and satisfaction under BPCI.**

BPCI was not associated with self-reported changes in functional status from before to after the anchor hospitalization, based on the beneficiary survey results. Differences in the rate of improvement and decline between BPCI and comparison respondents were small and not statistically significant for six out of seven measures for hospital-initiated episodes and five out of seven measures for PGP-initiated episodes. A smaller proportion of respondents treated by BPCI-participating hospitals reported favorable care experiences for six of nine measures ($p<0.05$) although the magnitude of the differences was small; three of nine measures for PGP-initiated episodes also indicated a small negative impact on care experience ($p<0.05$). BPCI respondents in hospital-initiated episodes were less likely than comparison respondents to report the highest levels of satisfaction with their overall recovery since leaving the hospital ($p<0.05$).

- **BPCI Model 3 had limited, mixed impact on quality of care.**

There are few indications that BPCI affected quality of care for Model 3 clinical episodes. The changes in claim-based quality measures were not statistically significantly different in most cases, and for those that were, the outcomes appeared to be due to the chance selection of particular comparison episodes. Assessment data indicated that beneficiaries with SNF-initiated orthopedic surgery episodes had less improvement in functional status, but this was partially due to shorter SNF stays for those with BPCI episodes. BPCI beneficiary survey respondents in surgical SNF-initiated episodes reported better mobility improvement than the comparison respondents, though BPCI MJRLE episode respondents reported less favorable care experience. BPCI respondents with

HHA-initiated MJRLE episodes were less satisfied with their care experience than comparison respondents and those with MJRLE or CHF episodes were more likely to report declines in functional status.

- **Reconciliation payments offset the reductions in payments due to BPCI, resulting in a net loss to Medicare of $202.1 million ($268 per episode) under Model 2 and $85.2 million ($921 per episode) under Model 3.**

The reductions in total episode payments did not translate into overall savings to Medicare after taking into account reconciliation payments, which reflect the performance payments to providers and other outlays under the initiative.[5] Using the impact estimates from the clinical episodes we analyzed, we extrapolated the total decline in payments due to BPCI to all Model 2 and 3 clinical episodes and participant types. After accounting for reconciliation payments through 2016, BPCI resulted in a net loss to Medicare of $202.1 million ($p<0.10$) under Model 2 and a net loss of $85.2 million ($p<0.10$) under Model 3. Reconciliation payments were greater than anticipated because CMS eliminated downside risk during part of the intervention.[6] If downside risk had not been eliminated, Medicare would have achieved a net savings of $144.3 million under Model 2 (not statistically significant) and a net loss of $55.2 million ($p<0.10$) under Model 3 through the end of 2016.

- **Because participation in Model 4 was low, it achieved little impact.**

There were no statistically significant changes in total payments and there were few statistically significant changes in utilization, quality of care, or patient satisfaction among the two clinical episodes analyzed through September 2015. (See the Year 3 Evaluation & Monitoring Annual Report referenced above for details about Model 4.)

- **Reductions in institutional PAC use and unplanned readmissions, as well as reduced patient complexity, were associated with greater per-episode NPRA under Model 2.**

The hospital and PGP EIs with the largest average standardized net payment reconciliation amount (NPRA) per episode relative to their target price (termed standardized NPRA) discharged a greater proportion of their patients home (with or without home health care) rather than to an institutional PAC setting and reduced unplanned readmissions relative to other participants. These changes may have been influenced by their circumstances at baseline. The hospital EIs with high standardized NPRA discharged a larger share of their patients to PAC during the baseline period than EIs with lower NPRA. Higher use of PAC in the baseline may have made it easier to lower PAC use under the intervention, which would generally translate into lower episode payments. EIs with relatively high standardized NPRA also experienced favorable changes in patient mix from the baseline to the intervention period. Less resource-intensive patients would be easier to discharge home without detrimental effects on quality than more complex patients.

[5] The reconciliation payments used for this analysis were extracted from the CMS accounting system.

[6] CMS eliminated the downside risk in the first five quarters for all participants because of concerns about the accuracy of the target prices. CMS later eliminated downside risk for PGP episodes because of episode attribution errors.

- **BPCI-participating providers differed from the average provider in ways that may have contributed to their ability or willingness to engage in the initiative.**

The provider and market characteristics of BPCI participants suggest that they may have had more resources to devote to responding to the initiative than providers that did not participate. The typical BPCI-participating hospital EI under Model 2 was a large, urban, non-profit facility with a teaching program. The majority of Model 2 PGPs were multi-specialty practices that increased their reliance on non-physician clinicians during the intervention period. BPCI-participating SNF EIs and HHA EIs under Model 3 were larger than the average non-participant and were more likely to be part of a chain.

BPCI participants also had higher payments prior to the start of BPCI than non-participants. In 2011, average Part A payments for BPCI-participating hospital EIs were $1,159 (6%) higher than average payments for non-participating hospitals (for clinical episodes with sufficient sample size for evaluation). Similarly, Model 3 BPCI-participating SNFs had payments that were $976 (4%) higher than payments for the same clinical episodes for non-participating SNFs. HHAs had payments that were $95 (1%) higher than non-participants. These data support our qualitative findings; participants we interviewed indicated that they selected the clinical episodes that presented the greatest opportunity to reduce payments.

- **The relative resource intensity of BPCI patients did not change during the intervention for Model 2 participants, except in two clinical episode strata.**

For the majority of Model 2 clinical episode strata there were no indications that the average resource intensity of patients changed from the baseline to the intervention period. For two planned, elective procedures, however, patients in BPCI episodes changed to a less resource-intensive mix in the intervention period, relative to the comparison group. For hospital-initiated non-cervical spinal fusion episodes and hospital- and PGP-initiated non-fracture MJRLE episodes, BPCI patients in the intervention period were less resource intensive than in the baseline, relative to the change in the comparison group. This is particularly notable because participants have the ability to identify patients for elective surgeries prior to admission and, therefore, could select less resource intensive patients to improve their ability to achieve positive NPRA. Qualitative data suggest another reason for the decline in patient resource intensity for these elective surgeries. Some participants said they postponed surgeries for higher risk patients until certain risk factors could be addressed to improve patient outcomes.

- **Evidence suggests that for some clinical episode strata, patient resource needs may have declined for SNF and HHA EIs relative to the change for comparison providers.**

In five SNF clinical episode strata and one HHA strata, patients treated by BPCI participants had better functional status upon admission during the intervention period than the baseline, relative to their respective comparison groups. Because SNFs and HHAs have considerable control over the patients they admit and they may evaluate patients prior to their admission decision, this may indicate that participants were choosing less intensive patients under the initiative.

- **Among Model 2 BPCI patients discharged to a SNF, the proportion who went to a SNF with a high star rating increased.**

The proportion of patients discharged to a SNF with a high star rating increased moderately from baseline to intervention across all six hospital-initiated clinical episodes and the one PGP-initiated clinical episode analyzed.

- **There was little evidence of market consolidation or concentration of PAC discharges under BPCI Model 2.**

We found that there was a small decline in median market share for BPCI participants from the baseline to the intervention period for six hospital clinical episodes (0.1 percentage points). For PGP EIs that enrolled in MJRLE in Q3 2015, there was a small increase in median market share during the same period (0.3 percentage points). BPCI participants under Model 2 did not decrease the number of SNF or HHA providers to which they discharged a majority of their patients in the intervention period relative to the baseline.

- **Use of program rule waivers, beneficiary incentives, and gainsharing was limited.**

BPCI participants could use beneficiary incentives to support their care redesign and use program rule waivers of certain Medicare requirements regarding telehealth, home visits, and the three-day inpatient hospital stay requirement for SNF coverage. Though many Awardees indicated they wanted to use these flexibilities, these waivers were not used in many episodes. The majority (82%) of Model 2 EIs were eligible to waive the three-day hospital stay rule because their Awardee agreements requested its use. However, only 35% of these EIs actually discharged a beneficiary to a SNF after a hospital stay of less than three days, and they did this for less than 5% of episodes with a SNF stay. Awardees cited difficulties in determining patients' BPCI eligibility and noted that the waiver became less important to their BPCI strategy as they reduced the share of patients receiving institutional PAC. Beneficiary incentives were the most widely selected option under Model 3, but fewer than 600 beneficiaries received an incentive for goods or services such as equipment, living accommodations, or transportation services.

Awardees could also disburse NPRA and internal cost savings (ICS) to partnering providers through gainsharing as a way to increase the engagement of their partners. Gainsharing plans were included in the agreements of 66% of Model 2 Awardees and 37% of these Awardees used the waiver to gainshare. By March 2017, they had distributed two-thirds of their NPRA and over a third of their ICS through gainsharing. Physicians, specifically orthopedic surgeons, were the most likely recipients of gainsharing under Model 2. Fewer Model 3 Awardees (18%) entered into gainsharing agreements. By March 2017, 71% of these Awardees distributed 21% of their NPRA to their gainsharing partners, mostly institutional PAC providers in the same BPCI convener group or other facilities in the same corporate system.

- **Although a large share of the intervention episodes were initiated by EIs that stopped participating in a given clinical episode, the relative declines in payments were robust to selective withdrawal.**

By the thirteenth quarter of the initiative, 83 hospital EIs (20%) and 71 PGP EIs (26%) ended their participation in all BPCI clinical episodes. The rate of attrition was similar under Model 3, in which 236 SNF EIs (27%) and 35 HHA EIs (30%) withdrew completely from the initiative.

The BPCI impact estimates do not include episodes initiated by providers after they withdrew from the initiative or stopped participating in a clinical episode. Because withdrawal from BPCI was non-random and was related to success in achieving NPRA, we estimated what the changes in payments would have been if these participants had not withdrawn. We found that our observed reductions in payments were generally robust and were not an artifact of providers with higher total payments leaving the initiative.

- **Model 2 hospitals were able to reduce episode payments for vulnerable beneficiaries without compromising quality of care.**

In four hospital clinical episodes, we examined the impact of BPCI on beneficiaries with characteristics that could make them especially vulnerable to changes in care, including those eligible for both Medicare and Medicaid, those with Alzheimer's disease or related dementias, and those who received institutional care within the five days prior to the anchor hospitalization. There was a relative decline in episode payments across all three vulnerable beneficiary populations in all four clinical episodes. Even with the reduction in payments, claim-based quality measures did not indicate worsening quality for BPCI beneficiaries in any vulnerable population across the four clinical episodes. While there were some cases where the survey responses indicated a statistically significant less favorable perception of care experience and satisfaction among these beneficiaries relative to the comparison group, these results did not carry over to survey outcomes of functional status or claim-based quality, which tended to either indicate improvement or no difference.

- **BPCI had no impact on total market volume of non-fracture MJRLE procedures.**

BPCI did not have a statistically significant impact on the volume of non-fracture MJRLE discharges within health care markets. There are concerns that an unintended consequence of bundled payments is that providers would generate additional episodes to generate additional NPRA. Such an increase in volume would offset reductions in Medicare payments. Our point estimates indicate that BPCI resulted in an average increase in non-fracture MJRLE discharges of 0.71% for the market. We also estimated an average increase of 2.14% in the volume of non-fracture MJRLE discharges among BPCI providers. However, those point estimates were not statistically significant ($p>0.20$).

C. Discussion and Conclusion

This fifth annual BPCI evaluation report presents results based on an average of six quarters of experience for both Model 2 and Model 3 participants. We estimated the impact of BPCI on select outcomes for 43 clinical episodes under Model 2 and 14 clinical episodes under Model 3.

Consistent with previous reports, results indicate that Medicare episode payments declined under the BPCI initiative. This conclusion is strengthened because the results remain robust even after accounting for the significant number of episodes that were initiated by participants that ultimately stopped participating in that clinical episode. Although non-random withdrawal could have positively biased the payment estimates, the results remain when all episodes, even from participants after they stopped participating in the clinical episode, are included. We continue to see general patterns of reduced intensity of PAC, with reductions in institutional care and increases in home health care. There are few indications in claims or beneficiary survey data that BPCI affected quality of care, either positively or negatively. Further, there is no evidence that the

changes in utilization due to BPCI resulted in decreased quality of care for vulnerable beneficiaries, as defined by dementia, dual eligibility, or prior institutional care.

Even though there is increased evidence that per episode payments declined, after considering reconciliation payments to participants, Medicare spending was higher under the initiative than it would have been absent of BPCI. However, if downside risk had not been eliminated for a portion of the intervention, Model 2 would have resulted in net savings to Medicare and Model 3 losses would have been lower.

Reconciliation payments to providers was one feature of this voluntary initiative that was intended to encourage providers and other entities to participate. Had the reconciliation payments been implemented differently to help ensure program savings, it is not known how participation in BPCI would have changed. The target price methodology, specifically the lack of risk adjustment and its basis on participant-specific historical payments, also may have reduced Medicare savings. Again, changes to the target price methodology to achieve savings would probably have affected participation.

There are limitations with this report. In particular, our primary analytic approach is dependent on how well the comparison group represents what would have happened absent the BPCI initiative. We matched providers and episodes on key factors that we expected to affect episode payments and quality. In most Model, episode initiator, and clinical episode combinations the matches were strong. For some combinations, the comparison episodes were not as close a match as we would like, even after multiple attempts to improve the match. In some instances, there were large differences in baseline levels of the outcome, which raises questions about whether the BPCI and matched comparison group had the same underlying trend in that outcome, a key assumption for the validity of the difference-in-differences (DiD). Sensitivity analyses, which included the unmatched BPCI and comparison episodes, also suggested that the statistical significance of some results may have been due to the chance selection of particular comparison episodes. As a result, our results for some outcomes among specific Model, episode initiator, and clinical episode combinations may be biased, although our overall conclusions that bundled payments has reduced episode payments while maintaining quality of care remains due to the consistency over time, across outcomes and clinical episodes.

The evaluation of the BPCI initiative is not complete. There are seven more quarters of claims and assessment data to evaluate. The additional data will allow more in-depth analyses of particular participant types, market effects, and beneficiary sub-populations. Additional research is needed to calibrate what type and level of financial rewards are required to entice participants into reducing episode payments without completely offsetting those lower payments.

I. Introduction

The Bundled Payments for Care Improvement (BPCI) initiative is designed to test whether linking payments for all providers that furnish Medicare-covered items and services during an episode of care related to an inpatient hospitalization can reduce Medicare expenditures while maintaining or improving quality of care. The Centers for Medicare & Medicaid Services (CMS) launched Phase 2 of Models 2, 3, and 4 of the BPCI initiative in 2013 under the authority of the Center for Medicare & Medicaid Innovation (CMMI).[7] BPCI Awardees, which may be hospitals, physician groups, post-acute care (PAC) providers, or other entities that convene health care organizations, entered into agreements with CMS to be held accountable for total Medicare episode payments. Those agreements also specified Awardees' choices among three payment Models, 48 clinical episodes, three episode lengths, and three risk tracks. Awardees also submitted BPCI implementation protocols that specified whether they would use available program rule waivers, beneficiary engagement incentives, or financial arrangements that could be protected under specific waivers of fraud and abuse laws.

The Lewin Group, with our partners Abt Associates, Inc., GDIT, Telligen, and Optum, is under contract to CMS to evaluate and monitor the impact of BPCI Models 2, 3, and 4. This report describes the impact of BPCI Models 2, 3, and 4 on Medicare payments, utilization, and quality through December 31, 2016.[8] This is the last annual report under this contract, so it incorporates results from all analyses.

A. BPCI Initiative

BPCI Awardees have incentives to reduce Medicare payments for the bundle of services in an episode of care. Under Models 2 and 3, Medicare episode payments are compared to a target price determined by CMS that is based on historical payments attributed to the episode-initiating provider for the same type of clinical episode. When aggregate Medicare episode payments are less than the target price, Awardees may receive net payment reconciliation amounts (NPRA) equal to the difference, which they can keep or share with their partnering providers. When aggregate episode payments are higher than the target price, Awardees may have to pay amounts to CMS. In this way, Model 2 and 3 Awardees have incentives to reduce aggregate episode payments. However, CMS eliminated downside risk for a portion of the initiative because of target price and episode attribution errors. Under Model 4, Medicare makes a prospective payment for the episode, so there is no reconciliation process. Awardees keep the difference between their costs and the prospective payment and receive no additional payment for costs above it.

Under each BPCI Model, an episode of care is triggered by a hospitalization for a Medicare Severity-Diagnosis Related Group (MS-DRG) contained in one of 48 clinical episodes (see **Appendix B** for a list of the 48 clinical episodes and associated MS-DRGs). The services provided during the clinical episode are bundled for payment purposes. Certain services, such as hospice,

[7] Model 1 began earlier than Models 2, 3, and 4 and concluded on December 31, 2016. Model 1 defined an episode of care as the inpatient stay in the acute care hospital. Under Model 1, Medicare paid hospitals discounted payments — based on the payment rates established under the Inpatient Prospective Payment System — but continued to pay physicians separately for their services under the Medicare Physician Fee Schedule.

[8] Due to small sample sizes, most outcomes were not examined for Model 4. See the Year 3 annual report for the most extensive discussion of Model 4 results.

readmissions for certain MS-DRGs, and some Part B services are excluded. The bundle definition and payment approach vary by Model as follows:

- **Model 2** has the most comprehensive bundle, which includes the triggering hospital stay (i.e., the anchor hospitalization) and all items and professional services (with certain exclusions) furnished within the chosen episode length of 30, 60, or 90 days post-discharge. The episode starts when a beneficiary is admitted to an episode-initiating acute care hospital or when the attending or operating physician for the beneficiary's hospitalization is in an episode-initiating physician group practice (PGP). Individual providers are paid regular Medicare fee-for-service amounts throughout the episode, and aggregate episode payments are reconciled retrospectively against the target price.
- The **Model 3** bundle includes items and services furnished after the anchor hospital discharge, within the chosen episode length of 30, 60, or 90 days. The episode starts when a beneficiary is admitted to an episode-initiating skilled nursing facility (SNF), home health agency (HHA), inpatient rehabilitation facility (IRF), or long-term care hospital (LTCH) within 30 days of discharge from a hospitalization for a chosen clinical episode. In the case of PGP episode initiators (EIs), the episode starts when a beneficiary is admitted to a PAC setting within 30 days of discharge from a hospitalization where the attending or operating physician for the beneficiary's hospitalization is associated with a participating PGP. Individual providers are paid Medicare fee-for-service amounts throughout the episode, and aggregate episode payments are reconciled retrospectively against the target price.
- The **Model 4** bundle includes the anchor hospitalization, all professional services during the anchor hospitalization, and any readmissions and associated professional services that occur within 30 days of discharge that are not explicitly excluded from the bundle. The admitting hospital is paid a prospectively determined amount and it, in turn, pays the providers furnishing services included in the episode. There is no NPRA for Model 4 because participants keep any difference between the prospectively determined amount and their payments to other providers for services furnished during the episode.

There are 336 possible unique combinations of Model, clinical episode, and EI provider type in BPCI across Models 2 and 3.[9] During the first 13 quarters of the initiative, episodes were initiated in 251 of the possible combinations. (See **Appendix C** for the count of episodes by Model and clinical episode during the first 13 quarters of the initiative.) Of these combinations, only 67 had sufficient participation and volume to support a regression-based difference-in-differences (DiD) analysis using a matched comparison group.[10] During the first 13 quarters of the initiative, 23

[9] In addition, Awardees may select one of three options for bundle length (30, 60, or 90 days) and risk track. There are three risk tracks that vary based on the exclusion of episodes with extreme payments from the reconciliation payment calculation based on winsorization. Risk track A includes episodes with payments that fall between the 1st and 99th percentile of national payments for that MS-DRG, risk track B includes the 5th to 95th percentile, and risk track C includes the 5th to 75th percentile.

[10] The results were stratified for major joint replacement for lower extremity (MJRLE) clinical episodes into fractures and non-fractures for Model 2 hospital and physician group practice (PGP) episode initiators and Model 3 SNF episode initiators. The results for coronary artery bypass graft (CABG) clinical episodes were

Model 4 hospital participants initiated episodes in 19 of the possible 48 clinical episodes. Eighteen of the 23 hospitals terminated their participation in BPCI by the end of Q4 2016. Due to the high rate of termination and the low volume of patient episodes, Model 4 was not evaluated in this report. See BPCI evaluation Year 2 report for Model 4 results.[11]

CMS implemented BPCI in a phased approach. Participants could apply for Phase 1, the preparation phase, and then transition into Phase 2, the risk bearing phase, over an extended period. The first participants began Phase 2 for at least some of their clinical episodes on October 1, 2013. By October 1, 2015, all participants had to transition their clinical episodes to Phase 2.

Please refer to the Year 3 annual report for additional detail on the BPCI initiative.[12]

B. Research Questions

This Annual Report provides a summative and formative evaluation of the BPCI initiative from October 1, 2013 through December 31, 2016 based on the evaluation and monitoring activities that the Lewin team completed across the five years of the contract. Three major research questions provided the framework for our analytic approach. More detailed questions have been added based on initial results and to meet CMS's evaluation needs. Each Model is considered separately under this evaluation.[13]

- What are the characteristics of the BPCI initiative and participants at baseline and how have they changed during the course of the initiative?

To understand initiative participants, their care redesign, model incentive structures, and use of waivers available under the initiative, we analyzed multiple data sources including Medicare claims, Provider of Service (POS) files, Area Health Resource Files (AHRF), other secondary sources, and data submitted by Awardees in their implementation protocols and quarterly data submissions. This information was supplemented by site visits, focus groups, and quarterly interviews with BPCI participants. This information provided context and explanatory variables to understand the impact of BPCI and the factors that contributed to the results of BPCI.

- What is the impact of the BPCI initiative on episode payments and the quality of care for Medicare beneficiaries?

We estimated the impact of BPCI on episode payments, utilization of services, and quality of care for Medicare beneficiaries, and provider referral patterns and market share. In this report, we also estimated the impact of BPCI on beneficiaries who may be more vulnerable to changes in service delivery or quality of care to determine whether BPCI had a disproportionate impact on these beneficiaries. Specifically, we examined three subpopulations of beneficiaries, those eligible for

stratified into emergent and non-emergent for Model 2 hospital episode initiators. The results for MJRLE fractures, MJRLE non-fractures, CABG emergent, and CABG non-emergent are included in the appendices.

[11] The report is available for download from: https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

[12] The report is available for download from: https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

[13] Please note that not all questions are addressed for all Models due to sample size. Participation in Model 4 was limited and declined over time. Therefore, we report few evaluation results for that Model. There was insufficient sample size in Model 3 and Model 4 to address questions about factors that contributed to the results of the initiative.

both Medicare and Medicaid, with Alzheimer's disease or related dementias, and receiving institutional care within five days prior to their anchor hospitalization admission.

Our main results estimate the impact of BPCI on payments based on the episodes initiated while the providers were still participating in the initiative. However, providers were able to stop participating in a BPCI clinical episode at any time, so these results may not reflect the full impact of the initiative. In this report, we also estimate the impact of BPCI on total episode payments assuming no providers stopped participating in a BPCI clinical episode to determine how withdrawals affected the estimated change in payments due to BPCI.

We have also added an analysis of the estimated impact of BPCI on Medicare program savings, which accounts for reconciliation payments. Our analysis combines the estimated changes in Medicare fee-for-service payments for clinical episodes due to BPCI with reconciliation payments to determine how the initiative affected overall Medicare spending.

BPCI-participating providers have incentives to increase their patient volume if they are receiving reconciliation payments. If participants generate additional episode volume under BPCI, Medicare spending will rise. We estimated the impact of BPCI on total market volume of non-fracture major joint replacement of the lower extremity (MJRLE) discharges. We focused on this clinical episode because it has the largest volume in BPCI and it is an elective procedure, so providers have more influence over patient decisions to have the surgery.

- What initiative, provider, beneficiary, and environmental factors contributed to the various results of the BPCI initiative?

There have been a range of responses to the BPCI initiative and a range of effects across participants. The analyses to address this research question are intended to identify the factors—either in participant characteristics or participant responses—that distinguish between those that achieved the initiative's objectives and those that did not. We investigated the initiative, provider, beneficiary, and environmental factors that contributed to BPCI initiative success or failure. We assigned EIs to performance categories based on their realized NPRA. We compared the top and bottom performers on their outcomes during the baseline period, changes in the composition of patient mix before and after BPCI implementation, and shifts in payment and utilization patterns before and after BPCI implementation.

C. Data Sources and Outcomes

This evaluation relied on multiple secondary and primary data sources to construct the sample, outcomes, and supplement the quantitative results with qualitative insights. We used provider-level data sources, including the CMS BPCI database, POS files, and Medicare Provider Enrollment, Chain, and Ownership System (PECOS) to identify and describe BPCI participant providers and select comparison providers. Medicare claims and enrollment data were used to construct episodes of care for patients at BPCI-participating sites (BPCI population) and at matched comparison providers. We also used claims and patient assessment data to create outcome measures and beneficiary risk factors associated with the outcomes.

We also collected primary data for this evaluation. We conducted multiple waves of a beneficiary survey to explore differences in patient care experiences and functional outcomes between

Medicare beneficiaries cared for by BPCI providers and similar beneficiaries whose providers do not participate in BPCI. We conducted over 100 site visits with BPCI-participating providers; focus groups to expand our understanding of the effect of BPCI on participants, their partners, and their markets; telephone interviews with key leaders of BPCI participants; and we convened five technical expert panels (TEPs). See **Appendix D** for more information on our secondary and primary data sources.

II. Model 2 Results and Discussion

A. Characteristics of the Initiative and Participants

This section summarizes the characteristics of the BPCI participants during the first 13 quarters of the initiative.

1. Key Findings

- 221 Awardees that represented 423 hospital EIs and 272 PGP EIs joined the risk-bearing phase of Model 2. Model 2 episodes accounted for 87% of the more than 796,000 episodes initiated across Models 2, 3, and 4 over the first 13 quarters of the initiative.
- BPCI-participating hospital EIs were larger and more likely to be non-profit, urban hospitals compared to non-participating hospitals. They also had 2011 standardized Part A payments that averaged 6% higher than non-participating hospitals.
- Of the PGPs included in the analysis, 20% were hospitalist practices, 26% were single-specialty practices, and 54% were multi-specialty groups. Most single-specialty practices were in the surgical specialty category, making up 18% of all PGPs.
- In BPCI-participating PGPs, up to 77% of clinicians were non-physician practitioners, and this share rose over time. PGPs also varied widely in their number of hospital discharges per quarter.
- The majority of BPCI participants requested through their applications the ability to use waivers from Medicare program requirements and fraud and abuse laws, but less than half of these Awardees reported using the flexibilities.
- During the first 13 quarters of the BPCI initiative, 20% of all Model 2 hospital EIs and 26% of all Model 2 PGP EIs withdrew completely from the initiative.

2. Methods

To characterize participants at the baseline and during the course of the initiative, Lewin conducted descriptive analyses using data from BPCI Awardee implementation protocols, Awardee-submitted data, POS files, CMS Inpatient Prospective Payment System (IPPS) annual files, Medicare claims, PECOS, and the AHRF. Lewin also analyzed qualitative data obtained through site visits, participant interviews, and focus groups. To specifically identify BPCI-participating PGPs, Lewin used data from the Tax Identification Number (TIN)/National Provider Identifier (NPI) crosswalk from the BPCI reconciliation contractor, Medicare claims data, and the CMS BPCI database. While 272 PGPs have participated in BPCI Model 2, the analysis includes only the 245 that existed in both the baseline and intervention period to follow the same cohort over time.[14]

See **Appendix D** for more information on variable definitions and for details of the qualitative methods.

[14] This was determined by the presence of at least one physician on the TIN/NPI crosswalk in each time period.

3. Results

a. Overall Model participant characteristics

Under Model 2, 221 Awardees that represented 423 hospital EIs and 272 PGP EIs participated in the risk-bearing phase of BPCI. Model 2 episodes accounted for 87% of the more than 796,000 episodes initiated in Models 2, 3, and 4 over the first 13 quarters of the initiative. Between Q4 2013 and Q4 2016, the average length of participation was seven quarters for hospital EIs and six quarters for PGP EIs. The majority of EIs (70% of hospital EIs and 92% of PGP EIs) joined BPCI in the last two quarters that they were able to enroll, namely Q2 and Q3 of 2015. As an indicator of the breadth of the BPCI initiative, 1,372 hospitals (28% of all acute care hospitals nationwide) either participated in BPCI as an EI or had a BPCI-participating PGP admitting patients between Q4 2013 and Q4 2016.[15]

The majority of hospital EIs reported having prior experience with care redesign and payment incentives (see **Appendix E** for the percentage of hospital and PGP EIs with prior experience in care redesign and payment incentives). PGP EIs were less likely than participating hospitals to report prior experience with care redesign and pay for performance incentives before joining BPCI. Furthermore, approximately 27% of BPCI hospital episodes and 25% of BPCI PGP episodes initiated through Q4 2016 was for a beneficiary who was aligned with a Medicare Accountable Care Organization (ACO).[16]

The average Model 2 EI, hospital or PGP, participated in eight clinical episodes; only three Model 2 EIs participated in all 48 clinical episodes. The most popular clinical episode among Model 2 EIs was MJRLE, in which 66% of all Model 2 EIs participated. Congestive heart failure (CHF) was the second most common clinical episode among hospital EIs, in which 43% of hospital EIs participated. However, the chronic obstructive pulmonary disease, bronchitis, asthma (COPD) episode was the second most common clinical episode among PGP EIs, with 36% that participated. The vast majority of episodes chosen by Model 2 hospitals and Model 2 PGPs were 90 days in length (97% and 99%, respectively). The count of EIs participating in each of the 48 clinical episodes is shown in **Appendix E**.

During the first 13 quarters of the initiative, a number of EIs stopped participating in some of their BPCI clinical episodes, but the majority remained in the initiative. In Model 2, 235 hospitals (56%) and 208 PGPs (76%) stopped participating in at least one clinical episode, and 83 hospitals (20%) and 71 PGPs (26%) withdrew completely from the initiative (see Section II.B.4 and II.C.3 for an analysis of the impact on payments of terminating participation among Model 2 hospital and PGP EIs, respectively).

b. Hospital characteristics

Compared with non-participating hospitals, BPCI-participating hospital EIs were more likely to be non-profit (77% vs. 57%), have a higher bed count (311 vs. 175), and have more discharges for BPCI episode MS-DRGs (3,004 vs. 1,598) in 2011, which was before BPCI was announced

[15] This figure includes discharges from Model 2 and Model 4 BPCI hospitals, as well as discharges made by attending and operating physicians that participated in Model 2 and Model 3 PGPs.

[16] Defined as participation in the Medicare Shared Savings Program, Pioneer ACO Model, or Next Generation ACO Model.

(Exhibits 1a and 1b). BPCI-participating hospitals also had larger teaching programs, as indicated by a higher resident-to-bed ratio (0.12 vs. 0. 05). Additionally, BPCI-participating hospital EIs were more likely to be located in urban areas (92% vs. 69%) than non-participating hospitals.

Interviews with Teaching Hospital Representatives and Rural Hospitals under BPCI

In Q2 2017, we conducted interviews with representatives from teaching hospitals under BPCI to learn more about their experiences. Interviewees noted a few advantages to being a teaching institution in their BPCI performance, including opportunities for research and having an abundance of staff. More specifically, the opportunities for research often allowed for the ability to share efforts and resources across hospital and PAC partners. Discussions with teaching hospitals also focused on resources that may be specific to their setting (e.g., increased access to advanced clinical care and resident programs). A few interviewees mentioned that their use of newer technologies, ability to perform more advanced procedures, or ability to work with more complex patients had positive effects, such as attracting new patients from outside of their catchment area and improving patient outcomes.

To better understand the perspectives of rural hospitals that participated in BPCI, we conducted interviews with rural hospital participants in Q2 2017. Interviewees noted that rural providers face greater challenges than urban ones in an initiative such as BPCI. These challenges included scarcity of potential partners (e.g., primary care physicians, specialists, and community services); lack of knowledge regarding programs such as BPCI among providers; limited internal staffing resources; limited patient access to transportation, which affected keeping follow-up appointments; and complex patient needs. Of note, Model 2 interviewees did not believe that being a rural provider affected their ability to collaborate with PAC providers, often because they had developed relationships or collaborative efforts prior to BPCI.

Participants and non-participants were similar with respect to disproportionate share percentage, a measure of their proportion of low income patients. Among BPCI participants, the average disproportionate share percentage was 27%, compared with 29% for non-participating hospitals. Additionally, the proportion of BPCI-participating hospital EIs that were part of a chain was comparable to that of non-participating hospitals (52% vs. 53%), and they had a similar share of total inpatient days attributed to Medicare beneficiaries (39% vs. 42%).

Exhibits 1a and 1b: Baseline Characteristics of BPCI-participating Hospital Episode Initiators and Non-participating Hospitals, Model 2

Domain	Characteristic	BPCI Hospital Episode Initiators (N)	BPCI Hospital Episode Initiators (%)	Non-participating Hospitals (N)	Non-participating Hospitals (%)
Ownership	For Profit	66	16%	638	23%
	Government	32	8%	542	20%
	Non-Profit	321	77%	1,594	57%
Urban/Rural	Rural	32	8%	872	31%
	Urban	387	92%	1,902	69%
Part of Chain	Yes	216	52%	1,469	53%

Characteristic	BPCI Hospital Episode Initiators (mean)	Non-participating Hospitals (mean)
Bed Count	311	175
Number of Discharges for BPCI Episode MS-DRGs, 2011	3,004	1,598
Medicare Days Percent	39%	42%
Resident-to-bed Ratio	0.12	0.05
Disproportionate Share Percent	27%	29%
Hospital Market Share	21%	27%

Note: Data from 419 BPCI hospital episode initiators and 2,774 non-participating hospitals. Medicare Severity-Diagnosis Related Group

Source: Lewin analysis of 2013 Provider of Service (POS) files and 2011 Medicare claims. BPCI-participating hospitals are defined as hospitals participating in Model 2. Non-participating hospitals are all other hospitals not participating in any BPCI initiative that reported values for all measures listed above and are not in Maryland. Please note that BPCI-participating hospitals that received Medicare certification after 2011 are not included in this table.

Notably, standardized Part A payments during the inpatient stay plus the 90-day post-discharge period (PDP) for patients discharged in BPCI MS-DRGs in 2011, which was before BPCI was announced, were higher for BPCI-participating hospital EIs than for non-participating hospitals (see **Appendix E** for 2011 standardized Part A payments for the inpatient stay plus 90-day PDP). Across all clinical episodes, average standardized payments were $1,159 (6%) higher among BPCI-participating hospitals than among non-participating hospitals. The difference in standardized payments varied by clinical episode; the greatest percentage difference in payments was for major joint replacement of the upper extremity episodes, where the mean payment was $2,766 (16%) higher in BPCI-participating hospital EIs than non-participating hospitals. The smallest difference was for acute myocardial infarction (AMI) where payments were $18 (0.1%) higher in BPCI-participating hospitals than in non-participating hospitals.

c. PGP characteristics

Using available data on BPCI-participating PGPs, we describe the distribution of physician and non-physician specialties, including the presence of hospitalist practices. We also consider average discharge rates as measures of practice size. While we describe the characteristics of BPCI-participating hospitals, skilled nursing facilities (SNFs), and home health agencies (HHAs) in 2011, prior to the initiative, we examine the characteristics of BPCI-participating PGPs from 2012 through 2016 because their physician composition is quite variable over time. Due to the lack of a comparison group, we cannot attribute these changes to the impact of participation in BPCI.

Clinician specialties

Physicians employed with BPCI-participating PGPs were associated with 58 different Medicare provider specialties. We grouped these physicians into seven categories that represented approximately 98% of physicians affiliated with BPCI-participating PGPs.[17] We defined primary

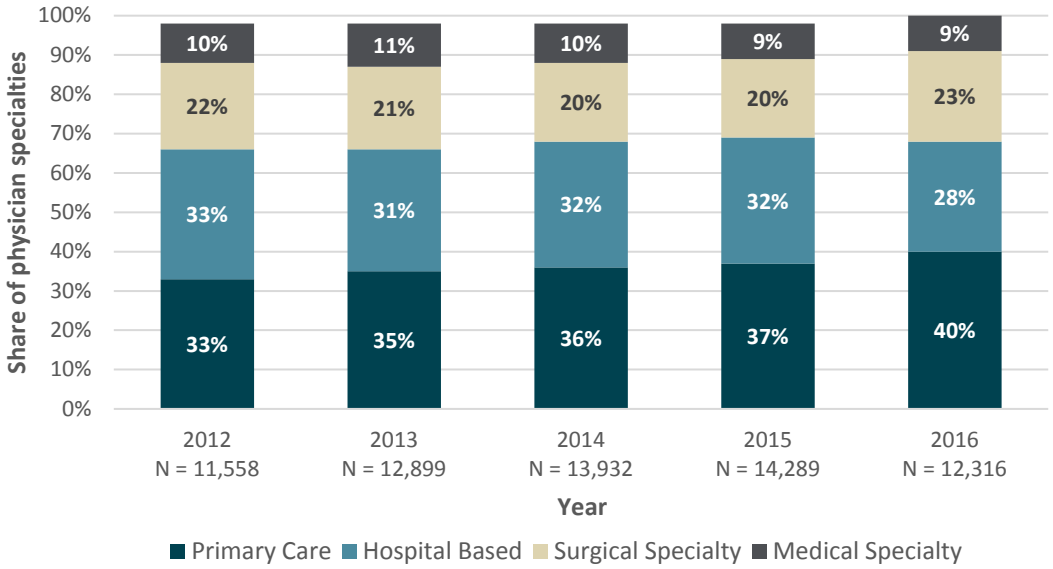
[17] We categorize clinician specialties following the methodology in Welch WP, Cuellar AE, Stearns SC, Bindman AB. (2013). Proportion of physicians in large group practices continued to grow in 2009–11. *Health Affairs* (Millwood). 32(9):1659-1666.

care physicians with more than 90% of their Part B charges incurred in the inpatient setting as hospitalists and other hospital-based specialties such as emergency medicine, critical care, and diagnostic radiology.

From 2012 through 2016, as shown in Exhibit 2, primary care physicians made up about one-third of all BPCI-participating physicians in 2012, increasing to 40% by 2016. The proportion of physicians in the other three categories fluctuated over the years, without a clear overall trend.

The total number of physicians associated with BPCI-participating PGPs increased from 11,558 in 2012, to 14,289 in 2015. It then declined to 12,316 in 2016 because of PGPs that withdrew from the initiative.

Exhibit 2: Distribution of Physician Specialties Among BPCI-participating PGPs, Model 2, 2012 – 2016



Note: This calculation counts physicians based on the length of time they were associated with a PGP during the calendar year. See **Appendix D** for definitions of specialty categories. Sample size varies by year due to the changing composition of PGPs on the TIN/NPI crosswalk and PGPs withdrawing from the intiative. PGP = physician group practice. TIN = tax identification number. NPI = national provider identifier. Physicians in psychiatry, obstetrics-gynecology, and other physician categories are excluded from this chart because they made up less than 2% of physicians associated with PGPs. Due to rounding, the percentages displayed in the graph may not add up to 100%.

***Source*:** Lewin analysis of BPCI TIN/NPI crosswalk as of Q4 2016; Medicare Part B claims, 2012 through 2016.

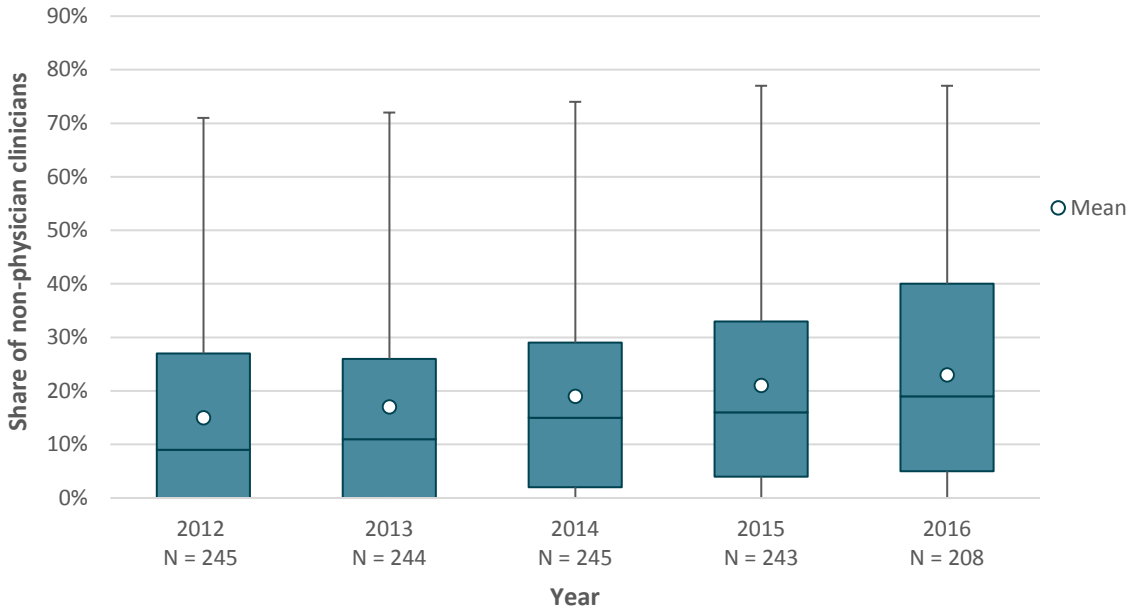
We classified PGPs into three categories based on the mix of physician specialties. PGPs were designated as hospitalist practices if 70% or more of their physicians were identified as hospitalists in any year.[18] The remaining PGPs were then categorized as either single- or multi-specialty. We classified BPCI-participating PGP as single-specialty if more than 90% of physicians were in the same specialty category in any year. Of the 245 PGPs included in the analysis, 50 (20%) as hospitalist practices, 63 (26%) were identified as single-specialty practices, while the remaining 132 (54%) were multi-specialty practices. Most single-specialty practices

[18] Based on the methodology described in Pete Welch W, Stearns SC, Cuellar AE, Bindman AB. (2014). *Use of Hospitalists by Medicare Beneficiaries: a National Picture.* Medicare Medicaid Res Rev. 2014; 4(2):mmrr2014.004.02.b01.

were in the surgical specialty category, which made up 18% of all PGPs. See **Appendix E** for further details about participant characteristics.

We also examined the employment of non-physician clinicians, such as nurse practitioners and physician assistants, at BPCI-participating PGPs. Nationally, studies have projected increased shares of non-physician clinicians in the clinician labor force.[19,20] BPCI-participating PGPs may use non-physician clinicians to perform certain services to reduce episode payments because they receive lower payments under the Medicare physician fee schedule. As shown in Exhibit 3, the proportion of non-physician clinicians at BPCI-participating PGPs grew from 2012 to 2016. The proportion of non-physician clinicians within practices varied greatly across BPCI-participating PGPs, ranging from 0 to 77% over all five years. The range, quartiles, and average of the proportions all increased over time. However, due to the lack of comparison PGPs, we were unable to determine whether this change was a result of BPCI participation.

Exhibit 3: Distribution of the Share of Non-Physician Clinicians within BPCI PGPs, Model 2, 2012–2016



Note: The minimum and maximum values are represented by the top and bottom bars. The blue box represents the 25th through 75th percentiles, and the horizontal line through the box represents the median. Box plots without minimum bars indicate that the minimum value and 25th percentile are zero. The share of non-physician clinicians is calculated as the number of non-physician clinicians divided by the total number of clinicans at the PGP. This calculation counts clinicians based on the length of time each individual was associated with a specific BPCI PGP during the calendar year. Sample size varies by year due to the exclusion of PGPs without NPIs on the TIN/NPI crosswalk for that year and PGPs that withdrew from the BPCI initiative. PGP = physician group practice. TIN = tax identification number. NPI = national provider identifier.

***Source*:** Lewin analysis of TIN/NPI crosswalk as of Q4 2016; Medicare Part B claims, 2012 through 2016.

[19] Bodenheimer R, Bauer L. Rethinking the primary care workforce – an expanded role for nurses. N Engl J Med 2016; 375: 1015-1017.

[20] Hooker RS, Brock DM, Cook ML. Characteristics of nurse practitioners and physician assistants in the United States. Journal of the American Association of Nurse Practitioners 2015; 28(1): 39-46.

Discharge rates by PGP

BPCI-participating PGPs varied from 0 to more than 10,000 average discharges per quarter for MS-DRGs included in the 48 BPCI clinical episodes (see **Appendix E** for the distribution of quarterly discharges, including mean, median, 25th percentile, and 75th percentile). While the overall range was large, the middle 50% of the distribution of the number of discharges per quarter was relatively compact in each year, falling approximately between 100 and 900. Additionally, variation existed in the quarterly discharge rate per physician at BPCI-participating PGPs, ranging from 0 discharges per quarter to more than 70.

We also compared the average number of quarterly MJRLE discharges for orthopedic surgeons at BPCI-participating PGPs versus all Medicare-billing orthopedic surgeons nationwide.[21,22] We found that the median number of MJRLE discharges per quarter from 2012 to 2016 ranged between eight and 10 for participating surgeons, compared to four or five discharges per quarter across all orthopedic surgeons who admitted Medicare beneficiaries for MJRLE. See **Appendix E** for detailed results.

d. Model incentive structure characteristics

Conveners in BPCI

Under BPCI, there are three types of conveners that vary based on whether they assume financial risk for the BPCI episodes initiated at their respective Awardees or EIs, and whether they joined the initiative under another organization. (See **Appendix A** for complete technical definitions of Awardee Conveners, Designated Awardee Conveners, and Facilitator Conveners.) Approximately 94% of Model 2 EIs participated under or as a convener. As described in prior evaluation reports, conveners of each type provide a range of services for their EIs. Exhibit 4 summarizes the key roles that conveners play in BPCI. Interviewees reported that conveners serve varying roles, including assisting with clinical episode selection, providing administrative support and data analysis, conducting care redesign activities, providing tools, and educating on best practices. These roles were previously described in the Year 3 BPCI Models 2-4 Evaluation Annual Report.[23]

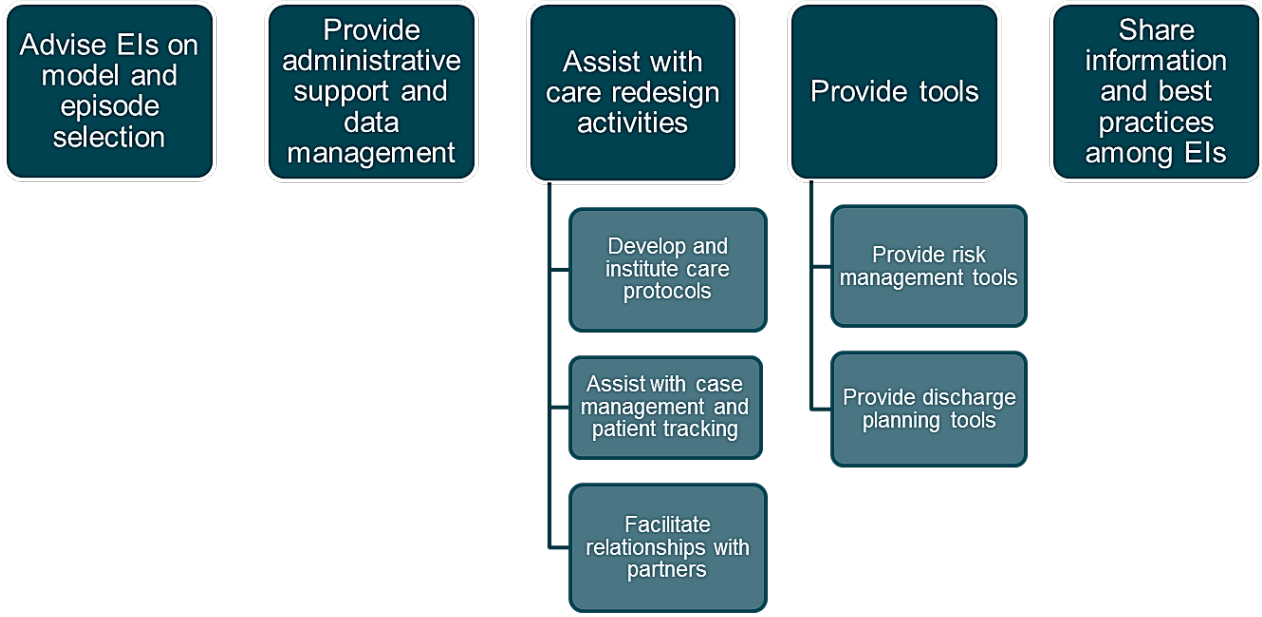
Interviewees also indicated that conveners may assist EIs with tracking patients after discharge through providing services such as case managers or call centers that conduct follow-up phone calls with patients. Other conveners offer EIs use of proprietary patient tracking tools and software. For example, one convener provides it's EIs with a software tool that provides an objective measure of a patient's functional and cognitive status and, based on a library of historical patient outcomes, predicts the patient's anticipated functional improvement at different PAC settings. Interviewees also described other systems that conveners provided to their EIs for predicting or tracking patient outcomes after discharge.

[21] For orthopedic surgeons with at least one MJRLE discharge in a given year.

[22] To calculate average quarterly discharges for all NPIs in the dataset, total discharges in a year were summed for each NPI that had at least one discharge in that year. The total was divided by four to obtain a quarterly average. This method was also applied to obtain the figures for MJRLE only, restricting to orthopedic surgeons. For BPCI physicians, this also included restricting to discharges that occurred during an active period on the TIN/NPI crosswalk.

[23] The report is available for download from: https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

Exhibit 4: Key Roles of Conveners in BPCI



Source: Lewin analysis of qualitative findings from site visits, focus groups, and participant interviews conducted with Model 2 participants between Q4 2013 and Q4 2016.

Waiver use

The BPCI initiative allows participants to use waivers from certain Medicare program requirements with respect to the three-day hospital stay, telehealth, and post-discharge home visits to facilitate the implementation of care redesign interventions. In addition, an Awardee may provide beneficiary incentives or engage in gainsharing agreements under BPCI that may be protected under the BPCI fraud and abuse law waivers. To be eligible to use the waivers from the three-day inpatient hospital stay requirement for SNF coverage or restrictions on beneficiary incentives and gainsharing, Awardees must describe the use of the waiver in their CMS-accepted implementation protocols. All participants have access to waivers regarding telehealth services and home visits.

A varying proportion of Model 2 EIs and Awardees were eligible to use the waivers, and eligibility did not necessarily coincide with reported use. Most Model 2 EIs (82%) were eligible to use the three-day hospital stay waiver because their Awardee requested the waiver in their accepted implementation protocol, but only 35% of those eligible used the waiver for 6,236 patients (4.4%) in BPCI episodes (Exhibit 5). Few EIs used the telehealth (11%) or post-discharge home visit (2%) waivers. More than half of Model 2 EIs (59%) were eligible to use the waiver for beneficiary incentives, and of those, 37% distributed incentives to 4,462 patients in BPCI episodes. Of the 146 Awardees (66%) that entered into gainsharing agreements, 54 (37%) distributed NPRA or internal cost savings (ICS) to their gainsharing partners.

Exhibit 5: BPCI Participation in Various Waivers, Model 2, Q4 2013 – Q4 2016

Waiver	Participants Allowed to Use Waiver (N)	Share of Participants Allowed to Use Waiver of all Participants (%)	Participants that Used Waiver (N)	Share of Participants that Used the waiver of those Allowed to Use Waiver (%)
Three-day hospital stay waiver	548	82	194	35
Telehealth	672	100	76	11
Post-Discharge Home Visit	672	100	13	2
Beneficiary Incentives	398	59	148	37
Gainsharing*	146	66	54	37

* The data for the gainsharing waiver are based on Awardees, while the three-day hospital stay waiver, beneficiary incentives, telehealth, and post-discharge home visit are based on episode initiators (EIs). Awardees must have had a CMS-accepted implementation protocol through Q3 2016 to be included in the gainsharing statistics, and the waiver use figures are based on Awardee-submitted data through Q2 2017. To be included in the other four waivers, EIs must have initiated at least one BPCI episode from Q4 2013 through Q4 2016.

Note: The 672 EIs that participated in Q4 2013 through Q4 2016 are distributed among 221 Model 2 Awardees. EIs include both hospitals and physician group practices (PGPs). The telehealth and post-discharge home visit waivers are available to all Model 2 EIs without specifying it in their Awardee's implementation protocol. Therefore, the denominator used to calculate percent of Model 2 EIs that used these two waivers is the total number of Model 2 EIs through Q4 2016.

Sources: Lewin analysis of Awardee implementation protocols for Q4 2016 BPCI participants, Medicare claims data for episodes initiated Q4 2013 through Q4 2016, and Awardee-submitted data Q4 2013 through Q4 2016. For gainsharing, Lewin analysis of Awardee implementation protocols for Q3 2016 BPCI participants and Awardee-submitted data Q4 2013 through Q2 2017.

Three-day hospital stay waiver

Similar to results from earlier reports, most site visit and participant interview respondents stated that they have not used the three-day hospital waiver for Medicare SNF coverage. We asked BPCI participants that could have used the waiver about barriers to waiver use. Many interviewees cited concerns about the Medicare coverage of the SNF stay, because of their inability to determine in real-time which patients were included in BPCI. For example, interviewees described cases where the 'working' MS-DRG was included in a BPCI clinical episode, but the revised discharge MS-DRG was not included in BPCI. In this situation, Medicare would not cover the SNF stay if the patient was discharged from the hospital after a stay of less than three days. A few interviewees noted that their convener recommended not using the waiver for this reason. In addition, a couple of interviewees indicated that because they were shifting patient discharge destinations from the SNF to home, they did not need the waiver as often as predicted. Some interviewees also felt that waiver use could result in unfavorable patient outcomes, such as readmission, by promoting inappropriately early discharges, or that it could increase the use of SNFs. Finally, a handful of participants believed that it was not clinically appropriate to discharge some patients (e.g., those with co-morbidities or non-surgical episodes) in less than three days, thus limiting the use of the waiver.

A handful of interviewees reported using the three-day hospital stay waiver. Most of these interviewees discussed using the waiver for certain clinical episodes, most commonly MJRLE. Limited availability of SNFs with a 3+ star rating in the market was noted as a challenge by interviewees that both did and did not use the waiver.

Beneficiary incentives

The waiver of certain fraud and abuse provisions allows EIs to provide beneficiaries with an in-kind, episode-related item or service that would not typically be allowed under existing law and regulation. From Q4 2013 through Q4 2016, 398 EIs (59%) were allowed to provide incentives because their Awardee included plans for beneficiary incentives in their implementation protocol (Exhibit 5, above). Of these Model 2 EIs, 148 EIs reported that they distributed 5,470 incentives to 4,462 patients in BPCI episodes (less than 1% of the total episodes initiated under Model 2 EIs). These incentives had an average cost that ranged in value from $12 to $817 per item or service (Exhibit 6). While the largest number of EIs were allowed to use home care and home visits incentives, transportation was the most common beneficiary incentive distributed, followed by medication management tools.

Exhibit 6: Beneficiary Incentives Distributed by Episode Initiators to Beneficiaries, Model 2, Q4 2013 – Q4 2016

Incentive description	Awardees Allowed to Use Incentives	EIs Allowed to Use Incentives and Had at Least One BPCI episode	Awardees that Provided One or More Incentives	Beneficiaries Receiving One or More Incentives	Average Cost per Incentive Provided
Transportation	26	120	10	1,675	$205
Equipment	18	290	10	1,368	$43
Home care/home visits	34	303	13	1,196	$817
Living arrangement services	12	72	7	257	$640
Telehealth/technology	16	234	5	68	$268
Wellness program/resources	13	278	2	90	
Medication management tools	14	228	2	1,586	$12

Note: A blank value indicates we could not calculate the average cost per incentive due to missing data. EI=episode initiator.

Source: Lewin Program Adherence Report based on analysis of Awardee-submitted data regarding disseminated beneficiary incentives through Q4 2016.

The beneficiary waiver was discussed in some interviews and site visits. One interviewee noted that they were considering providing telemonitoring services for high-risk patients through the waiver. A handful of interviewees reported that they were not using the waiver because they did not believe it was integral to their success in the initiative, because the administrative burden of tracking incentives outweighed the benefits of the waiver, or because they did not feel it would be useful for the clinical episodes they selected.

Telehealth and post-discharge home visit waivers

The telehealth and the post-discharge home visit waivers were rarely mentioned during site visits and participant interviews. However, we did hear reasons for not using the telehealth waiver including the requirement that patients receive telehealth services at specific sites of care (e.g., physician offices), the details of the waiver being too confusing, and the waiver not being useful for the clinical episodes in which the EI was participating. As for the post-discharge home visit waiver, one interviewee that was using the waiver noted that the waiver was difficult to understand and that it did not work well for the clinical episode they selected. Another interviewee stopped

using the post-discharge home visit waiver because they felt that the rules were too rigid, noting that there is "no way to right-size home health."

Gainsharing

About two thirds, or 146, of Model 2 Awardees had plans to gainshare. Among them, 48 reported distributing $144.4 million in NPRA and 19 reported distributing $10.6 million in ICS to their gainsharing partners between Q4 2013 and Q2 2017. These amounts represent about 67% of the total NPRA available for gainsharing through Q3 2016, and about 38% of the total available ICS realized through Q2 2017.

The most common type of gainsharing partner to receive NPRA and ICS distributions were individual physicians, followed by PGPs and hospitals (Exhibit 7). The majority of physicians that received a gainsharing distribution were orthopedic surgeons. On average, each gainsharing orthopedic surgeon received $23,005 in NPRA and $15,087 in ICS from Q4 2013 through Q2 2017.

Exhibit 7: Gainsharing Distributions Received by Partner Type, Model 2, Q4 2013 – Q2 2017

Gainsharing Partners	Awardees that Reported Eligible Partners of This Type	Number of Partners Receiving a Distribution	Number of Partners Receiving NPRA	Number of Partners Receiving ICS
Physicians	127	2,246	2,103	525
PGPs	64	172	168	9
Hospitals	63	71	66	10
Institutional PAC	19	26	26	0
HHAs	13	15	15	1
Other	16	12	12	0

Note: NPRA=net payment reconciliation amount. ICS=internal cost savings. PGP=physician group practice. PAC=post-acute care. HHA=home health agency.

Source: Lewin analysis of Awardee-submitted data collected February 2016 through August 2017 for Model 2 Awardees participating in BPCI between Q4 2013 through Q2 2017.

Similar to results in previous Annual Reports, the most common reason for using gainsharing was to increase engagement and improve collaboration with partners. Interviewees explained that they wanted to motivate specific changes in behavior, such as standardizing protocols and devices, achieving quality targets (e.g., length of stay, infections, readmissions, patient satisfaction), and decreasing or shifting PAC utilization. Most commonly, interviewees discussed gainsharing with physicians, though a couple mentioned gainsharing with PAC providers. Interviewees often reported that they did not gainshare with PAC providers because they believed that providing patient volume was a more significant incentive, or because gainsharing amounts would not be sufficient to compensate for lost revenue through foregone or shortened SNF stays.

Qualitative Findings Related to Challenges and Barriers to Participation in the Gainsharing Waiver

We conducted interviews and site visits with representatives from organizations that participated in gainsharing and asked them about their experiences. Those that participated in gainsharing mentioned two common challenges. First, interviewees noted that gainsharing funds could be difficult to distribute. For example, some interviewees cited challenges determining which physician "owned" a patient in a medical bundle. Second, interviewees discussed challenges with the gainsharing amount being subject to a cap set at 50% of the total Medicare fee-for-service (FFS) expenditures included in the episodes attributed to the physician. These interviewees noted that the limit on gainsharing dollars made it difficult to engage partners and motivate behavior change. One interviewee felt that, while the cap should not be eliminated, it should be raised to help address this challenge.

We also discussed gainsharing with some Model 2 interviewees whose organizations did not participate in gainsharing. Many of these interviewees did not believe that gainsharing would be impactful. These interviewees cited other factors that they felt were more powerful drivers of partner behavior change, such as providing volume for PAC providers. A couple of interviewees also mentioned that their physicians were highly engaged, making gainsharing unnecessary, or that gainsharing would not be useful for the episodes they selected. Other reasons for not gainsharing included direct physician employment and administrative or legal challenges to administering the program.

B. Impact of BPCI among Participating Hospitals

1. Payment, Utilization, Quality, and Patient Mix

This section presents the BPCI impact estimates on payments, utilization, quality, and the mix of patients for hospital-initiated episodes for the first 13 quarters of the initiative.

a. Key Findings

- There was a relative decline in total Medicare payments during the inpatient stay plus 90 days post discharge for 24 of the 32 Model 2 hospital clinical episodes for which we had sufficient sample size during the first 13 quarters of the BPCI initiative. Twelve of the 24 declines were statistically significant.
- SNF payments declined for nearly all of the clinical episodes. Smaller shares of patients were discharged to institutional PAC settings, and there were fewer SNF days for SNF users. These declines led to reduced Medicare payments.
- HHA payments increased, which is consistent with smaller shares of PAC users discharged to institutional PAC.
- In general, quality of care as measured through Medicare claims did not change under BPCI.
- Patient resource intensity did not change for the majority of clinical episodes from the baseline to the intervention period relative to a comparison group.

b. Methods

The analysis uses a DiD design to estimate the differential change in payment, utilization, and quality outcomes between the baseline and an intervention period for beneficiaries who received services from BPCI providers relative to beneficiaries who received services from a comparison group of non-BPCI providers. This approach controls for health care service use before the

hospitalization, beneficiary, market, and provider differences between BPCI and comparison episodes; eliminates biases from time invariant differences between the BPCI and comparison episodes; and controls for common trends in the BPCI and comparison population. We selected comparison providers and episodes to minimize differences in the distributions of characteristics between BPCI and comparison providers. Each BPCI hospital episode was randomly matched by quarter and MS-DRG to an episode from a comparison hospital.

The DiD estimates compare changes in outcomes from the baseline period (October 2011 through September 2012) to the intervention period (October 2013 through December 2016). Because participants may have started to implement changes in preparation for BPCI, we exclude Phase 1 of BPCI, the one-year period from October 2012 through September 2013. During this time, participants could begin signing up for BPCI but no participants had entered Phase 2, the risk-bearing or intervention phase.

We constructed comparison groups for 32 Model 2 hospital clinical episodes deemed to have a sufficient sample size for meaningful analysis. The episodes in these 32 clinical episodes represent 98% of all episodes initiated by Model 2 hospital EIs during the first 13 quarters of the initiative. A clinical episode was deemed to have sufficient sample size if there were 20 EIs with a total of 1,000 clinically relevant episodes, however, this minimum sample size was not based on a formal power calculation.

To assess whether statistically significant changes in the impact estimates could be due to the random selection of comparison episodes in the matched sample of providers, we also constructed DiD estimates that used all episodes from matched BPCI and comparison providers. We conducted this sensitivity test for the key quality and payment outcomes.[24]

See **Appendix D** for further details of the DiD methodology including data sources, outcome definitions, methods for identifying comparison populations, statistical models, tests for parallel trends between BPCI and comparison episodes in the baseline period, and details on the sensitivity test.

c. Results

Sample characteristics

We describe the BPCI Model 2 hospital analytical sample for the clinical episodes for which we had sufficient sample size in Exhibit 8. The number of BPCI hospital EIs included in the analysis ranged from 26 to 303 per clinical episode, and the number of episodes ranged from 1,218 to 109,786 episodes, depending on the clinical episode, from the beginning of BPCI in Q4 2013 through Q4 2016. Because providers could join BPCI over an extended period and could terminate participation in a clinical episode at any time or withdraw from the initiative completely, providers in this analysis participated on average six quarters from Q4 2013 to Q4 2016. During this period, 83 of 423 Model 2 hospital EIs (20%) withdrew entirely from the initiative. Among the 406 hospital EIs participating in any of the 32 clinical episodes analyzed in this report, approximately 54% stopped participating in at least one clinical episode by Q4 2016. The BPCI impact estimates

[24] The sensitivity test was conducted for total Medicare payments for the inpatient stay plus 90 days post discharge, all-cause mortality during the 30- and 90-day PDP, emergency department use during the 30- and 90-day PDP, and unplanned readmissions during the 30- and 90-day PDP.

include episodes from all hospital EIs during their participation period. The contribution of episodes from hospital EIs that stopped participating by Q4 2016 varies by clinical episode. For three clinical episodes, hospital EIs that stopped participating in the clinical episode had contributed 50% or more of the episodes during the intervention period. For an additional seven clinical episodes, hospital EIs that stopped participating contributed over 40% of the episodes. Section II.B.4 presents the analysis of the impact of withdrawal on payments, which suggests that relative payments continued to decline for BPCI participants in most clinical episodes relative to the comparison even after some EIs withdrew.

Exhibit 8: Characteristics of the Matched BPCI Providers Included in the BPCI Impact Estimates, Model 2 Hospitals, Q4 2013 – Q4 2016

Clinical Episode	BPCI Hospitals (#)	Matched Intervention Episodes (#)	Average Length of Participation (Quarters)	EIs that Terminated Participation in the Clinical Episode (#)	Episodes from EIs that Terminated (%)
Acute myocardial infarction	93	6,229	6	41	34.6
Cardiac arrhythmia	70	6,757	5	39	47.9
Cardiac valve	31	4,325	6	15	48.6
Cellulitis	79	6,034	6	36	47.3
Cervical spinal fusion	34	1,364	5	15	30.1
Congestive heart failure	173	37,330	6	57	28.5
COPD, bronchitis, asthma	133	21,438	6	44	31.0
Coronary artery bypass graft	43	3,622	7	14	33.6
Diabetes	45	1,698	6	19	31.0
Esophagitis, gastroenteritis and other digestive disorders	58	4,675	5	32	46.8
Fractures of the femur and hip or pelvis	47	1,245	6	17	33.0
Gastrointestinal hemorrhage	58	4,815	4	40	55.0
Gastrointestinal obstruction	51	1,936	5	29	46.0
Hip and femur procedures except major joint	101	8,514	6	38	25.0
Lower extremity and humerus procedure except hip, foot, femur	37	1,242	6	16	28.3
Major bowel procedure	46	3,415	6	20	33.7
Major joint replacement of the lower extremity	303	109,786	7	64	14.8
Major joint replacement of the upper extremity	26	1,540	6	9	24.9
Medical non-infectious orthopedic	94	7,512	6	41	33.7
Nutritional and metabolic disorders	57	3,022	5	31	51.6
Other respiratory	62	5,398	5	28	32.0
Other vascular surgery	36	1,732	6	18	39.4
Percutaneous coronary intervention	45	5,639	5	14	30.4

Clinical Episode	BPCI Hospitals (#)	Matched Intervention Episodes (#)	Average Length of Participation (Quarters)	EIs that Terminated Participation in the Clinical Episode (#)	Episodes from EIs that Terminated (%)
Renal failure	75	8,422	5	38	47.1
Revision of the hip or knee	32	1,261	6	16	33.9
Sepsis	119	29,888	6	55	39.0
Simple pneumonia and respiratory infections	132	25,423	6	42	24.9
Spinal fusion (non-cervical)	46	3,869	5	22	31.1
Stroke	77	13,216	6	30	20.8
Syncope and collapse	37	1,527	5	17	50.1
Transient ischemia	30	1,218	6	14	40.3
Urinary tract infection	83	9,192	6	32	27.4

Note: 10% of the BPCI hospitals were not included in the sample. This exhibit is limited to the BPCI providers used to calculate the DiD results in the remainder of this section. See **Appendix D** for information on the methods used to determine the sample. DiD=difference-in-differences. EI=episode initiator. COPD=chronic obstructive pulmonary disease.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2013 through Q4 2016 for BPCI providers.

Has patient mix changed under BPCI?

Because target prices are not risk adjusted, episode payments for BPCI participants could decline relative to the target price if their mix of patients during the intervention period was less resource intensive than their mix of patients during the baseline. Similarly, their episode payments could increase relative to the target price if their patient mix was more resource intensive in the intervention than in the baseline. To address the question of whether BPCI participant patient mix changed during the intervention, we examined claim-based patient characteristics that are associated with higher resource use. Exhibit 9 shows estimates of the change between the baseline and intervention period for BPCI patients relative to the change in the comparison group of patients for demographic characteristics, count of hierarchical conditions categories (HCCs– a risk measure used in Medicare's managed care program), and the utilization of care in the six months prior to the anchor hospitalization. For each of the measures in Exhibit 9, a negative value indicates a decline in the resource intensity of the BPCI patients during the intervention from the baseline period relative to the comparison group. Similarly, a positive value suggests a relative increase in patient resource intensity. (Please note: the impact analysis on payment, utilization, and quality presented below controlled for changes in these patient characteristics.)

We categorized Model 2 hospital strata (including MJRLE episodes stratified into fractures and non-fractures and coronary artery bypass graft (CABG) episodes stratified into emergent and non-emergent) into three broad groups: decline in patient resource intensity, increase in patient resource intensity, and no change.[25] Our categorization was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates. (See **Appendix D** for additional details of the categorization.)

Of the 34 Model 2 strata, the majority (26) did not have a consistent pattern of changes across measures. Five strata had indications that the BPCI-participating hospitals may have had a more resource intensive patient mix in the intervention period relative to the baseline period. Three strata had indications that BPCI patients were less resource intensive in the intervention period relative to the change for the comparison group. Two of these strata were planned, elective procedures (non-fracture MJRLE and non-cervical spinal fusion). This is particularly notable because participants have the ability to identify these patients prior to admission for these elective surgeries and, therefore, could select less resource intensive patients to improve their ability to receive reconciliation payments. Qualitative data suggest another reason for the decline in patient resource intensity for these elective surgery patients. Some participants said they postponed surgeries for higher risk patients until certain risk factors could be addressed, thus improving patient outcomes. (See **Appendix F** for additional details on which clinical episodes are planned, elective procedures.)

[25] The "no change" category includes strata that do not exhibit a consistent pattern toward a decline or an increase in patient resource intensity. This could be because they have indications of both decreases and increases in patient resource intensity or no statistically significant changes in either direction.

Exhibit 9: Relative Changes in Patient Resource Intensity, by Clinical Episode Strata, Baseline to Intervention, Model 2 Hospitals, Q4 2013 – Q4 2016

Change in Patient Mix	Clinical Episode Strata	Number of Episodes Q4 2013- Q4 2016	Age: 80+ years	Medicaid Eligibility	Disabled, no ESRD	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
Less resource intensive	Cardiac valve	4,325	-0.5	-0.9	-0.9	**-0.21**	-3.5	-0.9	-2.5	-0.3
	Major joint replacement of the lower extremity – Non-Fractures	95,186	-0.2	-0.4	**-0.6**	0.00	-0.3	-0.3	0.3	-0.3
	Spinal fusion (non-cervical)	3,869	-0.6	**-5.1**	**-5.4**	0.02	-1.4	**-3.7**	**-2.1**	-0.2
More resource intensive	Cellulitis	6,034	-1.3	2.1	0.9	0.03	**3.3**	**2.7**	**3.6**	-0.2
	Coronary artery bypass graft - Emergent	1,925	**5.8**	0.1	-0.6	**0.23**	2.0	3.4	**3.0**	**1.5**
	Medical non-infectious orthopedic	7,512	-0.5	**3.0**	**2.9**	0.02	0.7	0.6	-0.8	**2.1**
	Sepsis	29,888	-0.2	-0.5	0.3	**0.13**	1.2	0.6	**1.7**	0.0
	Urinary tract infection	9,192	-1.4	**3.1**	-0.3	0.09	1.4	1.7	**2.7**	-0.2
No consistent pattern	Acute myocardial infarction	6,229	1.5	0.7	-0.1	0.07	-0.6	-1.5	-1.0	0.8
	Cardiac arrhythmia	6,757	1.9	**2.6**	0.4	0.05	1.1	0.0	1.6	0.3
	Cervical spinal fusion	1,364	0.6	-3.6	-5.5	0.13	1.9	-0.1	-2.7	0.2
	COPD, bronchitis, asthma	21,438	-1.1	0.3	1.1	0.00	0.2	0.5	0.9	0.5
	Congestive heart failure	37,330	-0.2	0.3	0.2	0.02	-0.4	-0.9	0.9	0.1
	Coronary artery bypass graft - Non-Emergent	1,697	3.3	-2.1	-2.0	0.03	-1.9	-1.3	-0.9	-0.3
	Diabetes	1,698	3.2	-0.3	**-5.6**	0.11	2.0	**-8.3**	1.0	1.0
	Esophagitis, gastroenteritis and other digestive disorders	4,675	0.1	0.7	**-2.4**	-0.05	-1.1	1.1	-0.9	0.1
	Fractures of the femur and hip or pelvis	1,245	4.1	-3.7	0.3	-0.08	2.7	4.8	1.1	0.9
	Gastrointestinal hemorrhage	4,815	-2.1	1.6	0.9	0.04	2.5	1.8	1.5	0.3
	Gastrointestinal obstruction	1,936	-1.6	-0.3	1.7	-0.02	0.2	0.0	1.1	-1.4
	Hip and femur procedures except major joint	8,514	-0.8	1.1	-0.2	**0.11**	1.5	-1.2	1.0	1.0
	Lower extremity and humerus procedure except hip, foot, femur	1,242	**7.5**	-1.3	-5.1	0.19	4.7	-2.8	1.7	0.3
	Major bowel procedure	3,415	-0.7	1.3	-0.1	0.08	2.4	-0.3	2.8	0.7

Change in Patient Mix	Clinical Episode Strata	Number of Episodes Q4 2013- Q4 2016	Age: 80+ years	Medicaid Eligibility	Disabled, no ESRD	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
No consistent pattern (cont'd)	Major joint replacement of the lower extremity - Fractures	14,600	-0.4	-0.2	0.0	0.04	0.2	0.7	0.2	1.3
	Major joint replacement of the upper extremity	1,540	6.4	-4.7	-1.5	0.02	0.1	-3.0	0.8	1.5
	Nutritional and metabolic disorders	3,022	0.6	4.5	0.7	0.07	1.3	-0.7	3.8	0.8
	Other respiratory	5,398	0.7	-2.0	0.2	0.04	2.1	0.0	-1.2	-2.8
	Other vascular surgery	1,732	2.0	1.0	0.1	-0.08	0.2	3.1	2.0	-1.5
	Percutaneous coronary intervention	5,639	-0.9	0.7	-0.5	-0.08	-3.0	-1.0	-1.2	-0.7
	Renal failure	8,422	1.5	-1.1	-1.0	0.03	-0.1	-1.2	-1.6	1.2
	Revision of the hip or knee	1,261	1.9	-2.2	2.0	-0.01	0.5	1.1	1.5	-2.0
	Simple Pneumonia and respiratory infections	25,423	-0.2	-0.7	0.3	0.04	0.7	-0.5	0.6	-0.6
	Stroke	13,216	-0.8	0.0	-0.8	0.01	-0.3	-0.1	0.3	0.7
	Syncope and collapse	1,527	-0.9	-3.2	0.0	-0.14	0.0	-1.5	2.0	0.8
	Transient ischemia	1,218	-6.0	1.7	1.3	-0.06	-0.8	-1.5	1.6	-0.2

Note: Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Categorization of resource intensity was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates. See Appendix D for additional details of the categorization. ESRD=end-stage renal disease. HCC=hierarchical conditions categories. COPD=chronic obstructive pulmonary disease.

* These characteristics measure utilization of care in the six months prior to the anchor hospitalization. Count of HCCs is based on the six months prior to the anchor hospitalization.

Source: Lewin analysis of Medicare claims, enrollment data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q4 2016 (intervention period) for BPCI EIs and the matched comparison providers.

How have the average standardized payments changed under BPCI?

From Q4 2013 through Q4 2016, the total allowed payment amount (Medicare program payments plus coinsurance and copayments) for the inpatient stay plus the 90-day PDP declined from the baseline to the intervention period for the majority of BPCI hospital-initiated clinical episodes relative to the comparison group. Reduced SNF payments were the major contributor to these declines (Exhibits 10 & 11), which were often accompanied by concurrent increases in HHA payments (Exhibit 10). Detailed results of the BPCI impact estimates by clinical episode are located in **Appendix G**.

In 24 of the 32 clinical episodes we examined, there were declines in the total allowed payment amount for the inpatient stay plus 90-day PDP; the decline for 12 clinical episodes was statistically significant ($p<0.10$). Across these 12 clinical episodes, BPCI providers reduced total allowed payments by 4.4% on average relative to their payments absent BPCI (Exhibit 12).[26] The sensitivity tests suggest that the statistically significant decreases for three of these 12 clinical episodes may be due to the random selection of comparison episodes in the matched sample, and they may not have been statistically significant if a different sample of comparison episodes had been selected. (Results of the sensitivity test are presented with the BPCI impact estimates in **Appendix H**). The declines in total payments were due to reduced SNF payments, which declined for 29 of the 32 clinical episodes, and the decline was statistically significant for 16 ($p<0.10$). IRF payments decreased relative to the change in the comparison group for 15 of the 23 clinical episodes with sufficient sample size to examine; the decline in three clinical episodes was statistically significant ($p<0.10$). The HHA standardized allowed amount increased relative to the change in the comparison group in 28 of the 32 clinical episodes and the increase was statistically significant in 12 ($p<0.10$).

For the 10 clinical episodes that had the highest proportion of total baseline payments for PAC services, total allowed payments for the inpatient stay plus the 90-day PDP declined relative to the change in the comparison group in all but one. The decline in four of the high PAC payment clinical episodes was statistically significant ($p<0.05$). The SNF payments declined in nine of these clinical episodes, and the decline was statistically significant for seven ($p<0.10$). The HHA payments increased in all but one of these clinical episodes, and the increase was statistically significant for five ($p<0.10$). See the footnote for Exhibit 10 below or **Appendix F** for additional details on which clinical episodes have a larger proportion of total baseline payments for PAC services.

[26] BPCI providers' payments absent BPCI are calculated as what their payments would have been if they had experienced the same change in payments from the baseline to intervention period as the comparison group.

Exhibit 10: Impact of BPCI on Payment Outcomes, by Clinical Episode, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Total Amount Included in Bundle Definition[1]	Total Amount Paid by Medicare, IP through 90-day PDP[2]	Total Allowed Payment Amount, IP through 90-day PDP	SNF Standardized Allowed Amount, 90-day PDP[3]	HHA Standardized Allowed Amount, 90-day PDP[3]	IRF Standardized Allowed Amount, 90-day PDP[3]	Readmissions Standardized Allowed Amount, 90-day PDP[3]	Part B Standardized Allowed Amount, 90-day PDP	Total Part A & B 30-day Post-bundle
Acute myocardial infarction	6,229	-$394	-$363	-$431	-$263	$51	$49	$113	$69	-$262
Cardiac arrhythmia	6,757	-$654	-$549	-$653^	**-$358**	$61	-$70	-$206	-$107	-$224
Cardiac valve	4,325	-$522	-$348	-$374	-$280	$123	$27	-$342	**-$239**	-$259
Cellulitis[4]	6,034	-$404	-$532	-$664*	-$488	$102	-$52	-$120	-$134	**-$328**
Cervical spinal fusion	1,364	$363	$985	$866	-$347	$230	$505	-$208	-$14	-$353
Congestive heart failure	37,330	-$258	-$178	-$231	**-$260**	**$81**	-$62	-$136	$6	-$43
COPD, bronchitis, asthma	21,438	-$395	-$274	-$338	-$177	**$93**	-$43	$31	-$36	$60
Coronary artery bypass graft	3,622	-$1,342	-$835	-$907	**-$772**	-$39	-$553	$53	$45	$4
Diabetes	1,698	$470	$1,166	$763	**-$901**	$180		$482	-$94	-$148
Esophagitis, gastroenteritis and other digestive disorders	4,675	**-$1,015**	**-$1,098**	**-$1,265**	**-$698**	$43		-$378	-$131	-$285
Fractures of the femur and hip or pelvis[4]	1,245	-$597	-$788	-$978	-$594	$183	-$582	$269	-$215	-$347
Gastrointestinal hemorrhage	4,815	**-$929**	-$511	-$690	**-$586**	$87		$122	-$152	-$14
Gastrointestinal obstruction	1,936	$607	$875	$899	-$256	$115		$706	**$428**	-$249
Hip and femur procedures except major joint[4]	8,514	**-$1,848**	**-$1,381**	**-$1,857**	**-$2,020**	**$149**	$42	-$27	-$7	-$130
Lower extremity and humerus procedure except hip, foot, femur[4]	1,242	-$370	-$114	-$451	-$523	-$57	$90	$509	$96	-$649

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Total Amount Included in Bundle Definition[1]	Total Amount Paid by Medicare, IP through 90-day PDP[2]	Total Allowed Payment Amount, IP through 90-day PDP	SNF Standardized Allowed Amount, 90-day PDP[3]	HHA Standardized Allowed Amount, 90-day PDP[3]	IRF Standardized Allowed Amount, 90-day PDP[3]	Readmissions Standardized Allowed Amount, 90-day PDP[3]	Part B Standardized Allowed Amount, 90-day PDP	Total Part A & B 30-day Post-bundle
Major bowel procedure	3,415	-$976	-$1,222	-$1,491	-$577	**$260**	$3	-$121	-$160	-$290
Major joint replacement of the lower extremity[4]	109,786	**-$1,094**	**-$1,139**	**-$1,230**	**-$713**	$47	**-$411**	-$36	**-$56**	-$66
Major joint replacement of the upper extremity	1,540	-$20	$588	$607	$192	$250		$286	-$315	$51
Medical non-infectious orthopedic[4]	7,512	**-$1,644**	**-$1,317**	**-$1,689**	**-$1,365**	$108	-$215	-$222	-$101	-$215
Nutritional and metabolic disorders[4]	3,022	$179	$358*	$337*	-$744	$51		**$552**	$200	$444
Other respiratory	5,398	-$974	-$1,096	-$1,207^	-$407	$52	**-$360**	-$173	-$54	-$121
Other vascular surgery	1,732	-$375	$652	$603	$56	$11	$213	-$87	$230	-$168
Percutaneous coronary intervention	5,639	$352	$687*	$710*	-$9	$56	-$5	**$867**	-$76	$191
Renal failure	8,422	**-$1,134**	**-$803**	**-$995**	-$468	$47	-$146	-$126	-$41	-$238
Revision of the hip or knee	1,261	-$255	$417	$418	-$786	-$9		**$807**	$230	$307
Sepsis[4]	29,888	-$319	-$246	-$391	**-$440**	**$95**	$5	$19	-$18	-$144
Simple pneumonia and respiratory infections	25,423	**-$834**	**-$556**	**-$689**	**-$471**	**$61**	-$16	$38	-$27	-$10
Spinal fusion (non-cervical)	3,869	-$1,126	-$1,346	-$1,497^	-$482	-$116	-$429	-$342	**-$254**	$43
Stroke[4]	13,216	-$247	-$241	-$294	$15	$44	-$244	-$155	**-$160**	-$70
Syncope and collapse	1,527	-$1,021	-$1,149	-$1,346	-$641	$126		**-$698**	-$147	$10
Transient ischemia	1,218	**-$1,442**	**-$1,550**	**-$1,775**	**-$1,031**	$80		-$430	$100	-$192
Urinary tract infection[4]	9,192	**-$913**	**-$721***	**-$937***	**-$874**	**$144**	-$13	-$93	-$101	-$245

Note: The estimates in this table are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5%or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that

the outcome cannot be presented due to insufficient sample size. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. IP=inpatient. PDP=post-discharge period. SNF=skilled nursing facility. HHA=home health agency. IRF=Inpatient Rehabilitation Facility. COPD=chronic obstructive pulmonary disease.

[1] The total amount included in the bundle definition is based on only the 90-day episodes.

[2] Total amount paid does not include beneficiary out-of-pocket expenses.

[3] These payment measures are not conditional upon use of the service.

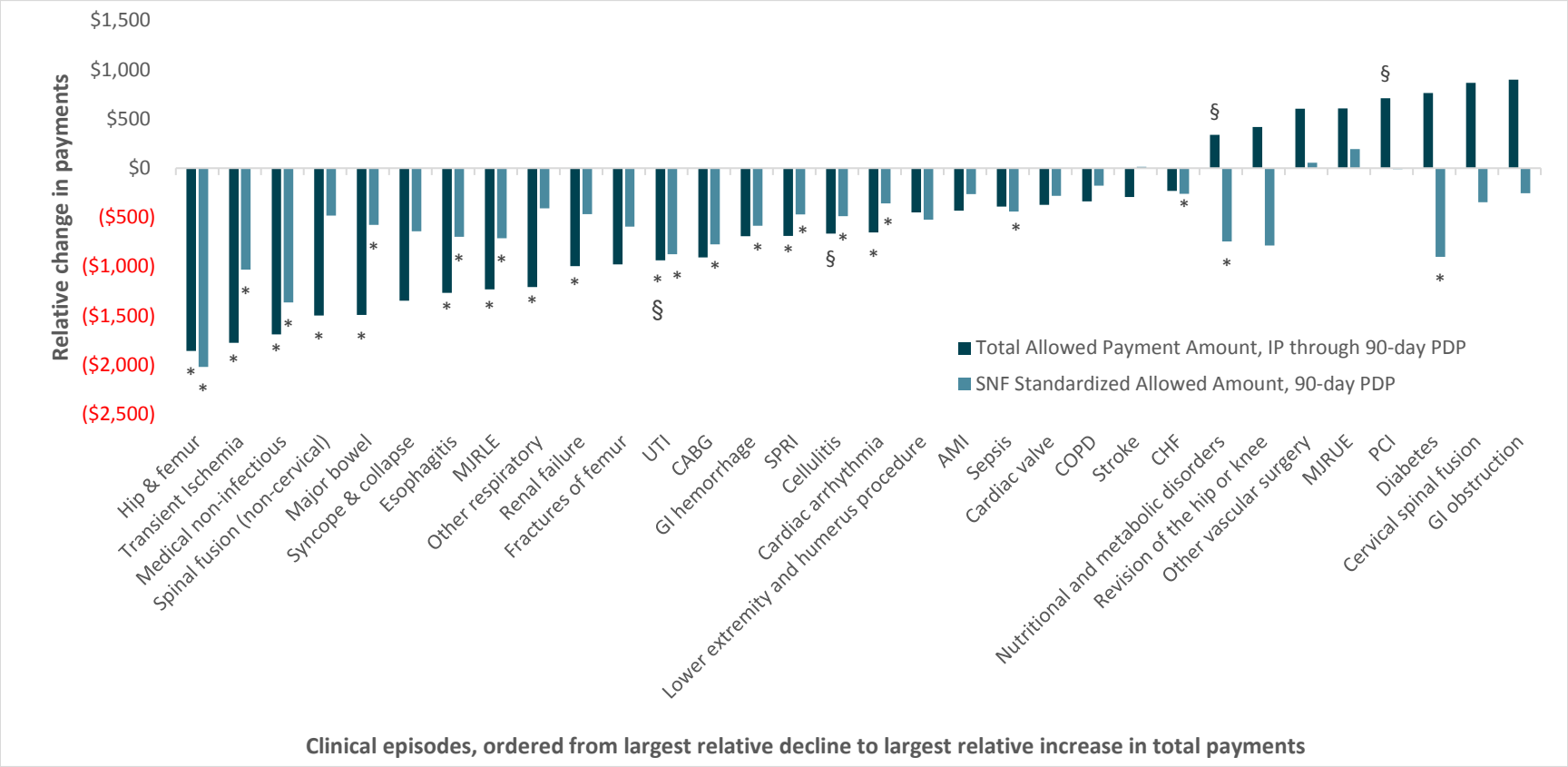
[4] This clinical episode is one of the 10 clinical episodes with the highest proportion of total baseline episode payments due to PAC payments.

***** This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate. Equal trends test was conducted for total allowed payment amount and total amount paid by Medicare, IP through 90-day PDP, emergency department visits, readmission, and mortality outcomes.

^ The sensitivity test results suggests that the statistically significant result may be due to the random selection of comparison episodes in the matched sample. See **Appendix D** for additional information on the sensitivity test methodology. Results of the sensitivity test are presented with the BPCI impact estimates in **Appendix H**.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 11: Impact of BPCI on SNF Payments and Total Medicare Allowed Payments, by Clinical Episode, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016



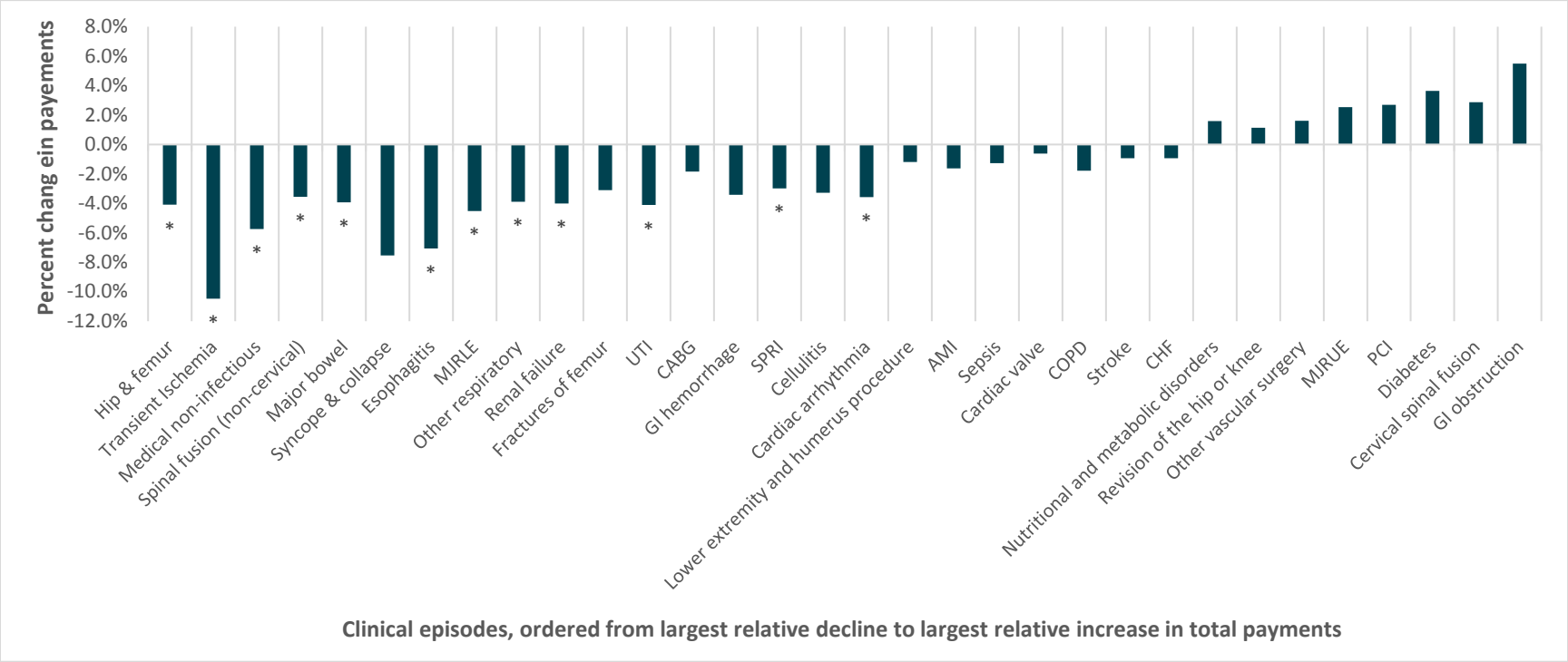
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and ordered starting with the largest relative decline in total allowed payment amount, IP-through 90-day PDP. These payment measures are not conditional upon the use of the service. SNF=skilled nursing facility. MJRLE=major joint replacement of the lower extremity. UTI=urinary tract infection. CABG=coronary artery bypass graft. GI=gastrointestinal. SPRI = simple pneumonia and respiratory infections. AMI = acute myocardial infection. COPD=chronic obstructive pulmonary disease. CHF=congestive heart failure. MJRUE=major joint replacement of the upper extremity. PCI=percutaneous coronary intervention.

*Indicates DiD estimates are statistically significant at the 10% level.

§ Data from the baseline period shows BPCI and matched comparison providers were not on parallel trends for this outcome, which is required for an unbiased estimate.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 12: Percent Change in BPCI Episode Allowed Payments from What Payments Would have Been Absent BPCI, by Clinical Episode, Model 2 Hospitals, Q4 2013 – Q4 2016



Note: The payments in this exhibit are the risk-adjusted standardized allowed amounts for the inpatient stay plus 90-day PDP and ordered starting with the largest relative decline. Episode payments absent BPCI, or the counterfactual, is the BPCI baseline payment amount plus the change in episode payment amount for the comparison group. The counterfactual can be expressed as: BPCI before + (Comparison after – Comparison before). The percent change can then be expressed as: (BPCI after – Counterfactual) / (Counterfactual). MJRLE=major joint replacement of the lower extremity. UTI=urinary tract infection. CABG=coronary artery bypass graft. GI=gastrointestinal. SPRI = simple pneumonia and respiratory infections. AMI = acute myocardial infection. COPD=chronic obstructive pulmonary disease. CHF=congestive heart failure. MJRUE=major joint replacement of the upper extremity. PCI=percutaneous coronary intervention.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

How have the services changed under BPCI?

Across all clinical episodes, there was no systematic change in the inpatient hospital length of stay (LOS) (Exhibit 13). Even among the nine clinical episodes with over 40% of their baseline payments attributable to the anchor stay, there was no clear pattern for changes in LOS. See the footnote for Exhibit 13 or **Appendix F** for additional details on which clinical episodes had over 40% of their baseline payments attributable to the anchor stay.

The changes in PAC service use were consistent with the changes in payments. For most clinical episodes, there was no statistically significant change in the proportion of patients discharged to PAC. Among patients who received any PAC, the proportion discharged to institutional PAC (SNF, IRF, or LTCH) declined in 26 of the 32 clinical episodes, and the decline was statistically significant for 13 clinical episodes ($p<0.10$) (Exhibits 13 and 14). Generally, the use of SNF care declined and the use of home health care increased (Exhibits 13 and 15). For BPCI episodes that used SNF care, the number of SNF days declined relative to the comparison group in 30 clinical episodes, and the decline was statistically significant in 17 clinical episodes ($p<0.10$). Although not directly related to payments, the number of home health visits for BPCI episodes increased relative to the comparison group in 27 clinical episodes, and the increase was statistically significant in seven ($p<0.10$).

There was a clear pattern with changes in PAC use in the 10 clinical episodes with the highest proportion of PAC payments in the baseline period. In all of these clinical episodes, the number of SNF days for SNF users declined relative to the comparison group, and the decline was statistically significant for six clinical episodes ($p<0.10$). Although an increase in HHA visits does not necessarily result in an increase in HHA payments, the number of HHA visits for HHA users increased for nine clinical episodes and was statistically significant for two clinical episodes ($p<0.10$).

Exhibit 13: Impact of BPCI on Utilization Outcomes, by Clinical Episode, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Anchor Hospital Stay LOS	Percent Discharged to PAC (pp)	Percent Discharged to an Institution out of Those who received any PAC[1] (pp)	Number of SNF Days, 90-day PDP[2]	Number of HHA Visits, 90-day PDP[3]
Acute myocardial infarction	6,229	0.00	-0.4	-0.2	-2.6	0.0
Cardiac arrhythmia	6,757	0.02	-0.2	-1.1	-2.4	1.5
Cardiac valve[5]	4,325	0.08	2.8	-9.5	1.0	-0.3
Cellulitis[4]	6,034	0.02	0.8	-1.1	-2.0	0.3
Cervical spinal fusion[5]	1,364	0.08	2.7	1.5	-3.0	0.5
Congestive heart failure	37,330	0.02	1.1	-0.9	-1.0	0.9
COPD, bronchitis, asthma	21,438	0.03	1.5	-1.2	-1.9	0.7
Coronary artery bypass graft[5]	3,622	-0.07	-4.4	-6.2	-2.4	0.6
Diabetes	1,698	0.09	-0.8	2.7	-9.4	-1.4
Esophagitis, gastroenteritis and other digestive disorders	4,675	-0.05	-0.3	-5.4	-3.8	1.3
Fractures of the femur and hip or pelvis[4]	1,245	0.07	0.2	-6.5	-0.6	1.1

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Anchor Hospital Stay LOS	Percent Discharged to PAC (pp)	Percent Discharged to an Institution out of Those who received any PAC[1] (pp)	Number of SNF Days, 90-day PDP[2]	Number of HHA Visits, 90-day PDP[3]
Gastrointestinal hemorrhage	4,815	-0.03	-2.2	-2.5	**-3.6**	0.3
Gastrointestinal obstruction	1,936	-0.08	1.1	-4.1	-0.1	**2.5**
Hip and femur procedures except major joint[4]	8,514	-0.03	-0.3	-0.6	**-3.8**	0.7
Lower extremity and humerus procedure except hip, foot, femur[4]	1,242	-0.09	-2.9	**8.1**	-4.4	-0.1
Major bowel procedure[5]	3,415	0.21	**4.4**	-4.1	-3.4	0.8
Major joint replacement of the lower extremity[45]	109,786	**-0.11**	**-3.1**	**-5.5**	**-2.2**	0.1
Major joint replacement of the upper extremity[5]	1,540	-0.01	4.8	1.1	-1.8	-0.1
Medical non-infectious orthopedic[4]	7,512	-0.05	-0.8	-0.9	**-4.2**	0.4
Nutritional and metabolic disorders[4]	3,022	-0.09	1.3	-4.6	-2.5	0.2
Other respiratory	5,398	0.00	-1.5	**-6.3**	-1.2	0.5
Other vascular surgery	1,732	-0.05	-0.9	0.8	-2.5	0.1
Percutaneous coronary intervention[5]	5,639	-0.04	-0.6	-2.2	-0.5	0.6
Renal failure	8,422	-0.02	0.2	0.0	**-2.6**	0.1
Revision of the hip or knee[5]	1,261	0.00	-0.7	**-11.8**	1.8	0.5
Sepsis[4]	29,888	0.10	0.7	-2.1	**-1.4**	0.0
Simple pneumonia and respiratory infections	25,423	-0.04	0.5	-1.8	**-2.7**	0.2
Spinal fusion (non-cervical)[5]	3,869	0.08	-4.7	-0.7	-1.5	**-1.1**
Stroke[4]	13,216	0.01	0.1	-0.4	-0.5	0.7
Syncope and collapse	1,527	-0.16	1.5	-3.1	**-6.6**	1.1
Transient ischemia	1,218	-0.02	-1.3	-7.7	-4.8	0.1
Urinary tract infection[4]	9,192	0.00	0.4	-2.7	**-2.7**	0.6

Note: The estimates in this table are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. PAC=post-acute care. PDP=post-discharge period. LOS=length of stay. pp=percentage points. HHA=home health agency. SNF=skilled nursing facility. COPD=chronic obstructive pulmonary disease.

[1]Institutional PAC includes SNF, IRF, and LTCH. Non-institutional PAC includes HHA.

[2]Beneficiaries must have spent a minimum of one day in a SNF setting during the 90-day PDP.

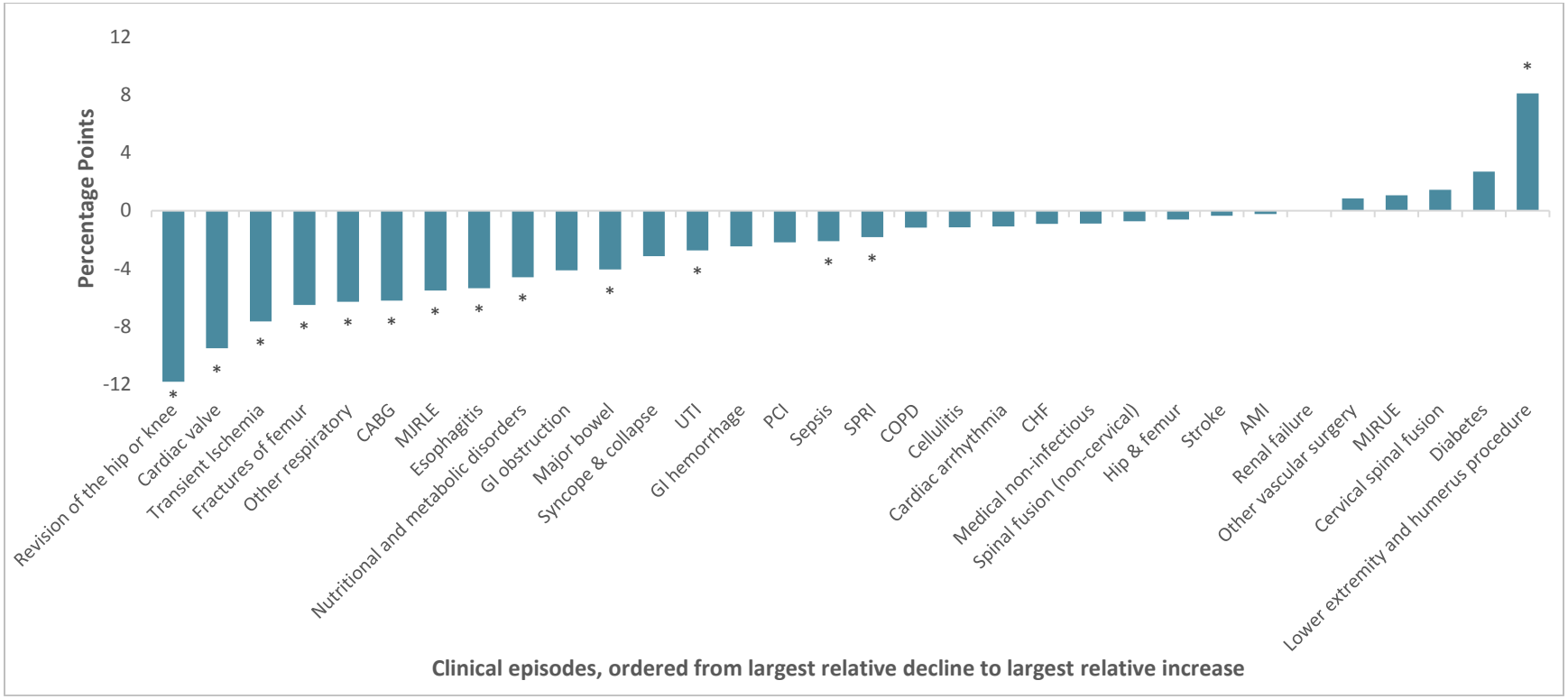
[3]Beneficiaries must have had a minimum of one HHA visit during the 90-day PDP.

[4]This clinical episode is one of the 10 with the highest proportion of total baseline episode payments due to PAC payments.

[5]This clinical episode is one of the nine with the highest proportion of total baseline episode payments due to the anchor inpatient stay (>40%).

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 14: Impact of BPCI on the Percent of Beneficiaries Discharged to Institutional PAC out of Those who Received any PAC, by Clinical Episode, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016



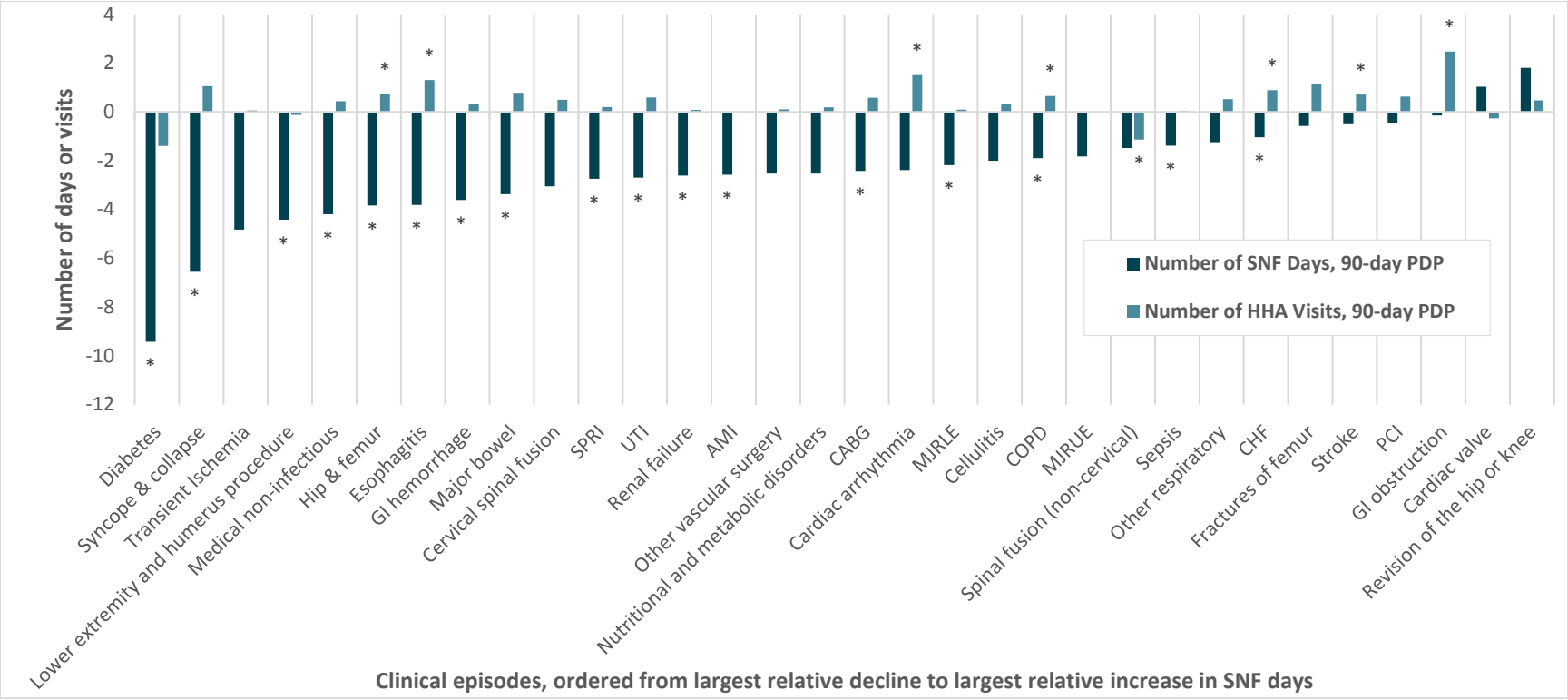
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model.

PAC=post-acute care. CABG=coronary artery bypass graft. MJRLE=major joint replacement of the lower extremity. UTI=urinary tract infection. GI=gastrointestinal. PCI=percutaneous coronary intervention. SPRI=simple pneumonia and respiratory infections. COPD=chronic obstructive pulmonary disease. CHF=congestive heart failure. AMI=acute myocardial infection. MJRUE=major joint replacement of the upper extremity.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 15: Impact of BPCI on SNF Days and HHA Visits, by Clinical Episode, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and ordered starting with the largest relative decline in the number of SNF days, 90-day PDP. These utilization measures are conditional upon use of the service. Beneficiaries must have spent a minimum of one day in a skilled nursing facility (SNF) and had at least one home health agency (HHA) visit during the 90 day post-discharge period (PDP) to be included in the DiD estimate for number of SNF days and HHA visits, respectively. SNF=skilled nursing facility. HHA=home health agency. GI=gastrointestinal. SPRI=simple pneumonia and respiratory infections. UTI=urinary tract infection. AMI=acute myocardial infection. CABG=coronary artery bypass graft. MJRLE=major joint replacement of the lower extremity. COPD=chronic obstructive pulmonary disease. MJRUE=major joint replacement of the upper extremity. CHF=congestive heart failure. PCI=percutaneous coronary intervention.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

How has quality of care changed under BPCI?

Results of the claim-based quality measures suggest that the quality of care generally did not change under BPCI Model 2 for hospital-initiated episodes (Exhibit 16). There were some statistically significant changes in individual quality outcomes, but most were not robust to the various sensitivity tests we conducted, as described below. Across clinical episodes, the predominant pattern was that there was no change in the measures, leading us to conclude there was no consistent impact on quality.

BPCI-participating providers had a relative decline in the mortality rate during the 90-day PDP in about half of the clinical episodes (19 out of 32); the decline was statistically significant for spinal fusion (non-cervical) ($p<0.05$). There was also a statistically significant increase in mortality for BPCI episodes in coronary artery bypass graft (CABG) clinical episodes, relative to the comparison group ($p<0.05$). This relative increase occurred in emergent CABG episodes. Since mortality is a relatively rare event and can be sensitive to random comparison episode selection, we conducted a sensitivity test in which we created an alternate set of DiD estimates that used all episodes from matched BPCI and comparison providers (see **Appendix D** for more information on the methods for the sensitivity analysis). The sensitivity test suggests that the statistically significant estimates for mortality may be due to the random selection of comparison episodes in the matched sample, and they may not have been statistically significant if a different sample of comparison episodes had been selected. (Results of the sensitivity test are presented with the BPCI impact estimates in **Appendix H**.) Furthermore, data from the baseline period shows BPCI and matched comparison providers were not on parallel trends for mortality for CABG episodes, which is required for an unbiased estimate. We also examined whether the relative increase in mortality for CABG episodes could be associated with increased hospice use, and we did not find evidence that this was the case.[27]

There was a statistically significant relative reduction in emergency department use during the 90-day PDP for three clinical episodes and a statistically significant relative increase for one ($p<0.10$). However, the sensitivity tests suggests that the statistically significant increase for transient ischemia episodes may be due to the random selection of comparison episodes in the matched sample, and it may not have been statistically significant if a different sample of comparison episodes had been selected (see **Appendix H** for results).

There were statistically significant relative declines in readmission rates for two clinical episodes ($p<0.10$). However, the results from the sensitivity test for both esophagitis and spinal fusion (non-cervical) suggest that the results may be due to the random comparison episode selection from matched providers, and they may not have been statistically significant if a different sample of comparison episodes had been selected (see **Appendix H** for results). In addition, data from the baseline period for esophagitis episodes shows BPCI and matched comparison providers were not on parallel trends, which is required for an unbiased estimate. There were statistically significant relative increases in readmission rates for four clinical episodes ($p<0.10$). The sensitivity test results suggest that the relative increases in readmission rates for fractures of the femur and hip or pelvis and revision of the hip or knee may be due to the random selection of comparison episodes, and they may not have been statistically significant if a different sample of comparison episodes

[27] The share of episodes with hospice use in the 90-day PDP was similar for BPCI and the comparison group, and controlling for it in the risk-adjustment regression model did not change the DiD results.

had been selected (see **Appendix H** for results). Furthermore, the relative increase in readmission rates for these two clinical episodes does not appear to be related to the decline in institutional PAC use.[28] We also examined whether the relative increases in readmissions could be associated with increased hospice use, and we did not find evidence that this was the case.[29]

Exhibit 16: Impact of BPCI on Claim-based Quality Outcomes, by Clinical Episode, Model 2 Hospitals, Baseline to Intervention, Q4 2013–Q4 2016

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	All-cause Mortality Rate, 90-day PDP (pp)	ED Use, 90-day PDP (pp)	Unplanned Readmission Rate, 90-day PDP (pp)
Acute myocardial infarction	6,229	-0.4	-0.5	1.3
Cardiac arrhythmia	6,757	0.0	1.3	-1.3
Cardiac valve	4,325	-0.8	0.7	-2.3
Cellulitis	6,034	0.2	-0.7	-1.2
Cervical spinal fusion	1,364	-0.1	-2.6	1.5
Congestive heart failure	37,330	0.2	-0.5	0.0
COPD, bronchitis, asthma	21,438	0.0	-0.4	-0.2
Coronary artery bypass graft	3,622	0.9*^	0.3	0.7
Diabetes	1,698	0.7	-0.2	-0.5
Esophagitis, gastroenteritis and other digestive disorders	4,675	-0.5	0.5	-2.3*^
Fractures of the femur and hip or pelvis	1,245	0.1	0.0	6.0^
Gastrointestinal hemorrhage	4,815	-0.2*	-0.7	-0.6
Gastrointestinal obstruction	1,936	-0.9	3.2	4.3
Hip and femur procedures except major joint	8,514	-0.6	1.3	-0.3
Lower extremity and humerus procedure except hip, foot, femur	1,242	-1.9	0.5*	3.4
Major bowel procedure	3,415	0.3	-0.6	0.9*
Major joint replacement of the lower extremity	109,786	-0.1	-0.1	-0.4
Major joint replacement of the upper extremity	1,540	0.6	-4.6	1.0
Medical non-infectious orthopedic	7,512	-0.6*	0.5	0.0
Nutritional and metabolic disorders	3,022	-0.1	0.5	2.6
Other respiratory	5,398	-0.3*	-1.3	-0.3
Other vascular surgery	1,732	-0.6	-3.4	-0.7
Percutaneous coronary intervention	5,639	0.2	-1.0	3.4*
Renal failure	8,422	0.7	-0.8	0.6
Revision of the hip or knee	1,261	0.3	2.1	3.7^
Sepsis	29,888	0.2	-0.2	-0.9
Simple pneumonia and respiratory infections	25,423	-0.2	0.1	-0.5*

[28] For episodes with a readmission, we examined the setting prior to the readmission (home without home health, home with home health, or institutional PAC), and there were no differences between BPCI and comparison episodes.

[29] The shares of episodes with hospice use in the 90-day PDP were similar for BPCI and the comparison group, and controlling for it in the risk-adjustment regression models did not change the DiD results.

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	All-cause Mortality Rate, 90-day PDP (pp)	ED Use, 90-day PDP (pp)	Unplanned Readmission Rate, 90-day PDP (pp)
Spinal fusion (non-cervical)	3,869	-0.6^	0.7	-1.9^
Stroke	13,216	-0.1	-1.7	-0.2
Syncope and collapse	1,527	-1.4*	-3.6*	-1.2
Transient ischemia	1,218	0.1	3.9^	-0.6*
Urinary tract infection	9,192	-0.5	1.2	-0.8*

Note: The estimates in this table are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. PDP=post-discharge period. pp=percentage points. ED=emergency department. COPD=chronic obstructive pulmonary disease.

* This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate. Equal trends test was conducted for total allowed payment amount and total amount paid by Medicare, IP through 90-day PDP, emergency department visits, readmission, and mortality outcomes.

^ The sensitivity test suggests that the statically significant result may be due to the random selection of comparison episodes in the matched sample. See **Appendix D** for additional information on the sensitivity test methodology. Results of the sensitivity test are presented with the BPCI impact estimates in **Appendix H**.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

2. Patient Functional Status and Health Care Experience

This section presents patient care experiences and functional outcomes from the beneficiary survey from May 2015 through June 2017.

a. Key Findings

- There were no consistent differences between BPCI and comparison respondents with regard to changes in self-reported functional status from before to after the anchor hospitalization.
- Relative to the comparison group, a smaller proportion of BPCI respondents reported favorable care experiences for six of nine measures. BPCI respondents were also less likely than comparison respondents to report the highest levels of satisfaction with their overall recovery since leaving the hospital.

b. Methods

We used a cross-sectional regression approach to estimate differences in self-reported outcomes between beneficiaries who were in Model 2 hospital-initiated BPCI episodes and a comparison group with the same clinical episode at non-BPCI participating hospitals. Survey outcomes included seven measures of self-reported change in functional status (improvement or decline) from before to after hospitalization, three measures of mental and physical health, nine measures of care experience, and a measure of overall satisfaction with physical recovery.[30]

[30] For measures of functional status, which include measures before and after the care episode, and measures of care experience and overall satisfaction with recovery, which pertain directly to the episode of care, we can confidently attribute estimated differences in those measures directly to the care episode. However, for measures

We sampled beneficiaries in 25 out of the 48 clinical episodes under Model 2 hospitals for the survey. BPCI and comparison beneficiaries were matched by age, presence of major complication or comorbidity (MCC), and hospital size. Clinical episodes with sufficient volume were also matched on hospital academic affiliation. Data were collected through seven waves of the survey, which began in May 2015 and ended in June 2017.[31] Responses from all clinical episodes and all waves were pooled for analysis.

Across all waves, the response rate averaged 47.3% and varied considerably by clinical episode, ranging from 75.4% to 35.0%. We applied non-response and sampling weights to all observations. Estimated differences between the BPCI and comparison respondents were risk-adjusted for beneficiary- and hospital-level characteristics. Because survey data were only collected during the intervention period, we can identify differences between BPCI and comparison respondents, but we cannot determine whether such differences were pre-existing or caused by BPCI.

See **Appendix D** for additional details about the survey measures, sample selection, and other methods.

c. Results

The survey results are based on 29,193 BPCI and 29,913 comparison responses.[32] The 25 clinical episodes included in the analysis represent 94% of all BPCI Model 2 hospital episode volume initiated since the start of the initiative.

Across all clinical episodes included in the survey sample, BPCI appears to have little association with self-reported changes in functional status from before to after the anchor hospitalization (Exhibit 17). Differences in the rate of improvement and decline were not statistically significant for six out of seven functional status measures, and point estimates were small and did not follow any consistent pattern. The exception was that BPCI respondents were 0.89 percentage points more likely to report increased use of a mobility device, relative to comparison respondents (BPCI 37.09 vs. 36.20 comparison; $p<0.10$).

The proportion of respondents with favorable care experiences was slightly smaller for BPCI survey respondents than comparison respondents. Relative to comparison respondents, BPCI respondents were less likely to report never receiving conflicting medical advice (-2.26 percentage points: 72.06 BPCI vs. 74.32 comparison); less likely to indicate that their level of care was always appropriate (-2.00 percentage points: 62.31 BPCI vs. 64.30 comparison); less likely to say they were discharged at the right time (-1.53 percentage points: 88.81 BPCI vs. 90.34 comparison); less likely to agree that medical staff took their preferences into account in deciding what health services they should have after leaving the hospital (-0.65 percentage points: 93.05 BPCI vs. 93.69 comparison); less likely to agree they had a good understanding of how to take care of themselves before going home (-0.91 percentage points: 94.71 BPCI vs. 95.62 comparison); and less likely to agree that medical staff clearly explained how to take their medications before going home (-0.90

of overall health, we cannot disentangle pre-existing differences in overall health from differences influenced by the episode. Therefore, we do not report overall health measures in the main body of the report. Risk-adjusted differences in overall health are reported in all survey results tables in **Appendix I**.

[31] For the exact months covered by each wave of the survey, see **Appendix D**.

[32] The difference is due to both a slightly higher response rate among the comparison group and some oversampling of the comparison group in order to increase statistical power for lower-volume clinical episodes.

percentage points: 93.90 BPCI vs. 94.80 comparison). BPCI respondents were also 1.90 percentage points less likely to indicate that they were "quite a bit" or "extremely" satisfied with their overall recovery since leaving the hospital than were comparison respondents (70.31 BPCI vs. 72.21 comparison; $p<0.01$).

Although BPCI respondents reported slightly less favorable care experience and overall satisfaction with recovery than comparison respondents, differences were small (ranging from -0.7 to -2.3 percentage points) and were not accompanied by worse functional status outcomes. In particular, BPCI respondents were equally likely to agree that they had been able to manage their health needs since returning home, which suggests that despite slightly more negative perceptions of care experiences, ultimately they were able to manage their health and experienced a similar recovery.

Exhibit 17: Differences in Survey-based Quality Outcomes between BPCI and Comparison Respondents, Model 2 Hospitals, May 2015 – June 2017

Domain	Survey Measure	BPCI Rate (%)	Comparison Rate (%)	Difference (pp)
Changes in Functional Status	Improvement in bathing, dressing, using toilet, or eating	72.94	72.78	0.15
	Decline in bathing, dressing, using toilet, or eating	14.60	14.87	-0.27
	Improvement in planning regular tasks	61.15	61.58	-0.43
	Decline in planning regular tasks	22.54	22.39	0.15
	Improvement in use of a mobility device (less likely to use)	49.79	50.53	-0.74
	Decline in use of a mobility device (more likely to use)	37.09	36.20	0.89
	Improvement in walking without rest	46.56	45.72	0.84
	Decline in walking without rest	26.56	26.85	-0.29
	Improvement in using stairs	45.88	45.72	0.16
	Decline in using stairs	30.06	29.81	0.25
	Physical/emotional problems limit social activities less frequently	60.74	60.99	-0.24
	Physical/emotional problems limit social activities more frequently	21.18	20.78	0.39
	Pain limits regular activities less frequently	60.41	59.86	0.55
	Pain limits regular activities more frequently	18.08	18.41	-0.33
Care Experience	Never received conflicting medical advice	72.06	74.32	**-2.26**
	Services always appropriate for level of care patient needed	62.31	64.30	**-2.00**
	Medical staff always spoke in patient's preferred language	92.97	93.56	-0.59
	Agree that patient was discharged at the right time	88.81	90.34	**-1.53**
	Agree that medical staff took patient's preferences into account in deciding post-discharge health care services	93.05	93.69	**-0.65**
	Agree that patient had good understanding of how to take care of self before going home	94.71	95.62	**-0.91**
	Agree that medical staff clearly explained how to take medications before going home	93.90	94.80	**-0.90**
	Agree that medical staff clearly explained what follow-up appointments would be needed before patient went home	94.58	95.05	-0.46
	Agree that patient had been able to manage health needs since returning home	96.38	96.16	0.22

Domain	Survey Measure	BPCI Rate (%)	Comparison Rate (%)	Difference (pp)
Overall Satisfaction	Extremely or quite a bit satisfied with overall satisfaction with recovery since leaving the hospital	70.31	72.21	-1.90

Notes: The estimates in this table are the result of a cross-sectional logistic regression risk adjustment model for binary indicators. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Estimates were based on 29,193 BPCI survey respondents and 29,913 comparison survey respondents. All responses were weighted for non-response and sampling design. The analysis pooled all 25 Model 2 hospital clinical episodes that were analyzed at the episode level, across survey waves 4-10. pp = percentage points.

***Source*:** Lewin analysis of BPCI Beneficiary Survey responses for episodes that began in May 2015 through June 2017.

Appendix I presents the results by clinical episode. The clinical episode-level results were generally consistent with the aggregate results. Roughly half of the estimates indicated no difference in care experience between BPCI and comparison respondents, while half indicated that a slightly smaller proportion of BPCI respondents reported having favorable care experiences. Similarly, in 18 out of 25 clinical episodes, there was no consistent difference between BPCI and comparison respondents in measures of self-reported functional status. Among the 7 clinical episodes where BPCI and comparison respondents indicated statistically significant differences across functional status outcomes, there were three clinical episodes where BPCI respondents indicated better outcomes and four where they indicated worse outcomes. This supports our conclusion that overall, BPCI was not associated with changes in functional status, although BPCI did have a small but negative association with care experience.

3. Impact on Payment and Quality among Vulnerable Beneficiaries

The previously discussed estimates of the impact of BPCI on payments, utilization, and quality, including beneficiary satisfaction, reflect the average experience across all beneficiaries in BPCI episodes. Although these estimates have shown no systematic evidence of reduced quality, it is possible that findings could differ for particular subpopulations of beneficiaries. We identified characteristics of beneficiaries who may be more susceptible to changes in service delivery or quality of care to investigate whether BPCI impacted them differently. We identified three subpopulations of beneficiaries with characteristics that could make them especially vulnerable to changes in care (henceforth "vulnerable populations"): those eligible for both Medicare and Medicaid, those with Alzheimer's disease or related dementias, and those who received institutional care within the five days prior to the anchor hospitalization. We conducted claim- and survey-based analyses for beneficiaries in four Model 2 hospital clinical episodes to answer two groups of research questions:

- How have the total Medicare payments and quality of care changed under BPCI among beneficiaries in vulnerable populations, compared to vulnerable populations in the comparison group?
- Are there differences in functional status, care experience, or satisfaction between BPCI beneficiaries and comparison beneficiaries with vulnerable characteristics?

a. Key Findings

There was no evidence to suggest that the impact of BPCI was systematically different for beneficiaries in the vulnerable populations examined in this analysis relative to similar beneficiaries in the comparison group. Differences in payments and quality results were not

consistent across types of vulnerable populations or outcomes, though we did observe some notable findings in individual clinical episodes or vulnerable populations:

- Reductions in episode payment for beneficiaries in vulnerable populations relative to similar beneficiaries in the comparison group were generally in line with the relative reductions in payments for all beneficiaries.
- Among BPCI beneficiaries with dementia in the MJRLE clinical episode, there was evidence of substitution of emergency department visits for unplanned inpatient readmissions.
- Claim-based quality measures did not indicate declines in quality for BPCI beneficiaries relative to similar beneficiaries in the comparison group in any vulnerable population across the four clinical episodes analyzed.
- In some cases, survey responses indicated a less favorable perception of care experience and overall satisfaction among vulnerable BPCI beneficiaries compared with similar beneficiaries in the comparison group, particularly in the CHF and sepsis clinical episodes. However, this perception did not appear to affect self-reported functional status or claim-based quality results, which tended to either indicate relative improvements or were not statistically significant.

b. Methods

We identified three characteristics that indicate beneficiaries who may be more vulnerable to changes in service delivery. These characteristics are Medicare and Medicaid dual-eligibility status, a diagnosis of Alzheimer's disease or related dementias, and institutional care use within the five days prior to the anchor hospitalization. Because beneficiaries may have more than one type of vulnerability, these categories are not mutually exclusive. Additional details on data sources and definitions for these categories are located in **Appendix D**. We examined the impact of BPCI on these subpopulations for four hospital clinical episodes under Model 2: MJRLE, CHF, sepsis, and simple pneumonia and respiratory infections (SPRI). These clinical episodes were selected based on high patient volume, which was necessary to ensure a large enough sample for sufficient statistical power, and to represent variation in clinical characteristics (e.g., acute versus chronic clinical episodes, medical versus surgical clinical episodes).

We included four claim-based outcomes in our analysis: total allowed payment amount for the inpatient stay plus the 90-day PDP, the 90-day post-discharge emergency department use rate, the 90-day unplanned hospital readmission rate, and the 90-day post-discharge all-cause mortality rate. We also examined seven measures from the beneficiary survey that assessed functional status, care experience, and overall patient satisfaction: improvement in need of a mobility device, improvement in walking without rest, improvement in using stairs, never received conflicting medical advice, discharged at the right time, understand care of self, and satisfaction with recovery. These survey measures were selected because they had shown notable findings in other aspects of the BPCI evaluation or because they could indicate potential quality concerns for a vulnerable population.

Separate approaches were used to analyze the claim-based and survey-based outcomes. For the claim-based outcomes, we used the DiD approach to measure the risk-adjusted difference for BPCI patients with each vulnerable characteristic relative to this difference for comparison patients with

the same characteristic. This approach seeks to identify the unique impact of BPCI on outcomes for vulnerable patients treated by BPCI providers after adjusting for other patient, provider, and market characteristics. For the survey outcomes, we developed risk-adjusted cross-sectional estimates of the difference between BPCI and comparison respondents with each vulnerable characteristic because there is no pre-BPCI survey data. Additional details on the analytical models and other methods can be found in **Appendix D**.

c. Results

Study sample

Exhibit 18 displays the total number of BPCI patient episodes and the proportion with the vulnerable number in each vulnerability category used in the analysis of claim-based outcomes for the four clinical episodes. The number of episodes identified as having one of the three vulnerable characteristics ranged from 2,177 to 11,949 across clinical episodes, and the prevalence of vulnerable characteristics ranged from 2% to 40%. Due to the high prevalence of dementia in some clinical episodes, we compared the prevalence of dementia across age categories with prevalence rates published in peer-reviewed literature. The prevalence of dementia in our BPCI samples was generally similar to or slightly greater than that of the overall Medicare population. The full results of this analysis are located in **Appendix J**.

Exhibit 19 shows the total number of BPCI survey respondents and the number in each vulnerable population for each of the four clinical episodes. Compared to the claim-based analysis, the number of BPCI survey respondents identified as having one of the three vulnerable characteristics was much smaller. Sample sizes for beneficiaries with vulnerable characteristics ranged from 141 to 650 respondents, which accounted for 6% to 26% of all BPCI survey respondents within the given clinical episode. Due to insufficient sample size, we were unable to conduct the survey analysis for beneficiaries with recent institutional use in the MJRLE clinical episode. Additional sample sizes for each BPCI and comparison group subpopulation are in **Appendix J**.

Exhibit 18: Prevalence of Vulnerable Characteristics in Four Clinical Episodes, Model 2 BPCI Hospitals, Q4 2013 – Q4 2016

Population	MJRLE N (%)	CHF N (%)	Sepsis N (%)	SPRI N (%)
Total episodes	109,786 (100%)	37,330 (100%)	29,888 (100%)	25,423 (100%)
Dual-eligible	11,445 (10%)	7,926 (21%)	9,554 (32%)	6,429 (25%)
Dementia	10,913 (10%)	10,459 (28%)	11,949 (40%)	9,365 (37%)
Recent institutional use	2,177 (2%)	3,748 (10%)	6,692 (22%)	3,753 (15%)

Note: MJRLE=major joint replacement of the lower extremity. CHF=congestive heart failure. SPRI=simple pneumonia and respiratory infection.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2013 through Q4 2016 for BPCI providers.

Exhibit 19: Prevalence of Vulnerable Characteristics for Four Clinical Episodes among BPCI Survey Respondents, Model 2 Hospitals, October 2014 – June 2017

Population	MJRLE N (%)	CHF N (%)	Sepsis N (%)	SPRI N (%)
Total BPCI survey respondents	3,175 (100%)	2,493 (100%)	2,518 (100%)	2,321 (100%)
Dual-eligible	216 (7%)	428 (17%)	551 (22%)	397 (17%)
Dementia	186 (6%)	511 (20%)	650 (26%)	607 (26%)
Recent institutional use		155 (6%)	246 (10%)	141 (6%)

Note: A blank cell indicates the vulnerable population was excluded from the analysis due to insufficient sample size. MJRLE=major joint replacement of the lower extremity. CHF=congestive heart failure. SPRI=simple pneumonia and respiratory infection.

***Source*:** Lewin analysis of BPCI beneficiary survey data for MJRLE episodes that began in October 2014 through June 2017 and for CHF, sepsis, and SPRI episodes that began in May 2015 through June 2017.

Major joint replacement of the lower extremity

Claim-based outcomes for the MJRLE clinical episode are shown in Exhibit 20. Overall, there were several statistically significant findings noted for the vulnerable populations in BPCI compared to their respective comparison groups, but there was no evidence to suggest a worsening of care across the vulnerable populations.

Among all three vulnerable populations, there were statistically significant declines in total allowed payments for beneficiaries in MJRLE episodes relative to beneficiaries with the same vulnerability in the comparison group. The overall results also showed a statistically significant relative decline for this measure, indicating that cost savings may still be achievable for vulnerable beneficiaries.

Although there were no statistically significant differences in mortality, four of the six estimates on readmissions and emergency department use were statistically significant. Dual-eligible beneficiaries in BPCI had a statistically significant ($p<0.10$) reduction in emergency department use during the 90-day PDP compared to dual-eligible beneficiaries in the comparison group. Beneficiaries with recent institutional use had a statistically significant ($p<0.10$) relative decline in unplanned readmissions. We also identified that BPCI beneficiaries with dementia had a statistically significant increase in emergency department use ($p<0.05$) in the 90-day PDP, as well as a statistically significant ($p<0.05$) decrease in unplanned readmissions. The similar magnitude and opposite directions of these changes suggest that a proportion of patients with dementia who would have been re-admitted in the absence of BPCI may have instead been evaluated in the emergency department without being admitted.

Exhibit 20: Impact of BPCI on Major Joint Replacement of the Lower Extremity Payments and Quality of Care, by Vulnerable Population Category, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016

Population	Total Allowed Payment Amount, IP through 90-day PDP	ED Use, 90-day PDP (pp)	Unplanned Readmission Rate, 90-day PDP (pp)	All-cause Mortality Rate, 90-day PDP (pp)
All beneficiaries	**-$1,230**	-0.1	-0.4	-0.1
Dual-eligible	**-$1,178**	-1.6	-0.7	-0.3
Dementia	**-$2,021**	**2.0**	**-2.4**	-0.2
Recent institutional use	-$2,002	-0.1	-4.0	-2.9

Note: The estimates in this table are results of a difference-in-differences (DiD) model. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. IP=inpatient. PDP=post-discharge period. ED=emergency department. pp=percentage points.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 21 includes the survey-based results for vulnerable beneficiaries in the MJRLE episode. Among all MJRLE episodes, four of the seven measures analyzed here had statistically significant differences between BPCI and comparison beneficiaries that suggested potentially worse results among BPCI beneficiaries. However, among the vulnerable subpopulations, there were only two measures that had statistically significant differences. Dual-eligible BPCI beneficiaries reported that they were less likely to never receive conflicting medical advice than dual-eligible beneficiaries in the comparison group ($p<0.10$). Although the magnitude is relatively large, the power of this test was low, with a 90% confidence interval ranging from -17.1 to -1.1. We also observed a statistically significant improvement ($p<0.05$) among BPCI beneficiaries with dementia in the use of stairs. The other functional status estimates were mostly positive for both vulnerable populations, indicating improvement, but they were not statistically significant. The magnitudes and directions of the remaining estimates for the vulnerable populations do not demonstrate a clear trend, providing little evidence of negative effects on vulnerable subpopulations. However, sample sizes are small, leading to relatively low statistical power.

Exhibit 21: Impact of BPCI on Major Joint Replacement of the Lower Extremity Survey Measures, by Vulnerable Population, Model 2 Hospitals, October 2014 – June 2017

Population	Functional status: Improvement in use of mobility device (less likely to use) (pp)	Functional status: Improvement in walking without rest (pp)	Functional status: Improvement in using stairs (pp)	Care experience: Never received conflicting medical advice (pp)	Care experience: Discharged at right time[1] (pp)	Care experience: Understand care of self before going home[1] (pp)	Satisfaction with recovery[2] (pp)
All beneficiaries	-2.3	1.8	-0.1	**-2.9**	**-2.1**	-0.8	-2.0
Dual-eligible	-3.5	3.2	1.8	-9.1	-0.9	0.7	3.7
Dementia	6.6	5.2	**12.4**	-0.9	-1.3	-0.2	1.6

Note: The estimates in this table are the results of a cross-sectional logistic regression risk adjustment model for binary indicators. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Recent institutional use was excluded from the analysis due to insufficient responses from beneficiaries that had recent institutional use. pp=percentage points.

[1] Measure reflects that respondents either agreed or strongly agreed with this statement.

[2] Measure reflects that respondents were either quite a bit satisfied or extremely satisfied with their recovery since leaving the hospital.

Source: Lewin analysis of beneficiary survey data for episodes that began October 2014 through June 2017 for BPCI and comparison hospitals.

Congestive heart failure

Claim-based results for vulnerable beneficiaries in the CHF clinical episode are shown in Exhibit 22. Across each vulnerable population and outcome, there were no statistically significant differences between BPCI and comparison episodes.

Exhibit 22: Impact of BPCI on Congestive Heart Failure Payments and Quality of Care, by Vulnerable Population, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016

Population	Total Allowed Payment Amount, IP through 90-day PDP	ED Use, 90-day PDP (pp)	Unplanned Readmission Rate, 90-day PDP (pp)	All-cause Mortality Rate, 90-day PDP (pp)
All beneficiaries	-$231	-0.5	0.0	0.2
Dual-eligible	-$458	-0.7	-0.5	-0.9
Dementia	-$179	-0.7	0.0	1.1
Recent institutional use	$277	-1.2	0.7	2.0

Note: The estimates in this table are results of a difference-in-differences (DiD) model. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. IP = inpatient. PDP=post-discharge period. ED=emergency department. pp=percentage points

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Survey-based outcomes for vulnerable populations in the CHF clinical episode are shown in Exhibit 23. In the functional status domain, BPCI beneficiaries with dementia reported a statistically significant improvement ($p<0.05$) in the use of a mobility device compared to patients with dementia in the comparison group.

In the care experience domain, BPCI beneficiaries with recent institutional care were less likely to never report receiving conflicting medical advice ($p<0.05$) and lower satisfaction with recovery ($p<0.05$) relative to similar beneficiaries in the comparison group. Both of these differences were quite large. BPCI dual-eligible beneficiaries were less likely to report being discharged at the right time ($p<0.10$). In addition, BPCI beneficiaries with dementia reported lower satisfaction with recovery than similar beneficiaries in the comparison group ($p<0.10$).

While these results indicate a less favorable perception of care experience for these populations, they did not appear to translate into reductions in functional status or quality as measured by claim-based outcomes. The direction of the estimates for the functional status survey results all indicate improved functioning for these two vulnerable populations, though, as mentioned above, only one estimate was statistically significant ($p<0.05$). In addition, none of the claim-based DiD estimates showed a statistically significant worsening among the vulnerable populations, and the estimates had mixed directions and magnitudes.

Exhibit 23: Impact of BPCI on Congestive Heart Failure Survey Measures, by Vulnerable Population, Model 2 Hospitals, May 2015 – June 2017

Population	Functional status: Improvement in use of mobility device (less likely to use) (pp)	Functional status: Improvement in walking without rest (pp)	Functional status: Improvement in using stairs (pp)	Care experience: Never received conflicting medical advice (pp)	Care experience: Discharged at right time[1] (pp)	Care experience: Understand care of self before going home[1] (pp)	Satisfaction with recovery[2] (pp)
All beneficiaries	0.0	-0.7	1.0	-1.1	-2.1	-0.8	-2.3
Dual-eligible	0.5	-4.1	3.7	0.9	-4.2	-1.1	-2.0
Dementia	**5.5**	3.0	1.6	-3.1	0.3	-0.6	-6.6
Recent institutional use	8.1	4.7	2.0	**-17.1**	-3.3	3.1	**-13.9**

Note: The estimates in this table are the results of a cross-sectional logistic regression risk adjustment model for binary indicators. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. pp=percentage points.

[1] Measure reflects that respondents either agreed or strongly agreed with this statement.

[2] Measure reflects that respondents were either quite a bit satisfied or extremely satisfied with their recovery since leaving the hospital.

Source: Lewin analysis of beneficiary survey data for episodes that began May 2015 through June 2017 for BPCI and comparison hospitals.

Sepsis

Exhibit 24 shows the claim-based results for the sepsis clinical episode. Relative to the comparison group, there were no statistically significant changes in total allowed payments, emergency department use, unplanned readmissions, or mortality among BPCI beneficiaries in the sepsis episode.

Exhibit 24: Impact of BPCI on Sepsis Payments and Quality of Care, by Vulnerable Population, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016

Population	Total Allowed Payment Amount, IP through 90-day PDP	ED Use, 90-day PDP (pp)	Unplanned Readmission Rate, 90-day PDP (pp)	All-cause Mortality Rate, 90-day PDP (pp)
All beneficiaries	-$391	-0.2	-0.9	0.2
Dual-eligible	-$587	-0.8	0.8	0.4
Dementia	-$538	-0.4	0.4	0.2
Recent institutional use	-$759	-0.9	0.2	1.8

Note: The estimates in this table are results of a difference-in-differences (DiD) model. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. IP = inpatient. PDP=post-discharge period. ED=emergency department. pp=percentage points.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

For beneficiaries in the sepsis clinical episode, survey results suggest less favorable care experiences and satisfaction, though findings were not consistent across vulnerable populations or outcomes (Exhibit 25). There were no statistically significant findings for measures in the functional status domain for any of the subpopulations, and there were no statistically significant differences in outcomes between BPCI and comparison beneficiaries with recent institutional use.

The proportion of dual-eligible BPCI respondents with favorable perceptions of care experiences and overall patient satisfaction was smaller than comparison respondents. BPCI dual-eligible respondents were less likely to report that they never received conflicting medical advice ($p<0.05$), less likely to agree that they had a good understanding of how to take care of themselves before going home ($p<0.05$), and less likely to report that they were extremely or quite a bit satisfied with recovery since leaving the hospital ($p<0.05$). However, these perceptions did not appear to affect self-reported functional status or claim-based quality results. All three functional status measures were positive, and claim-based measures were small in magnitude; no results were statistically significant.

The share of BPCI survey respondents with dementia who reported that they never received conflicting medical advice was statistically significantly smaller than comparison beneficiaries with dementia ($p<0.05$). BPCI beneficiaries with dementia were also less likely to report being discharged at the right time ($p<0.10$). Results for the remaining survey- and claim-based outcomes were both negative and positive in direction and none were statistically significant.

Exhibit 25: Impact of BPCI on Sepsis Survey Measures, by Vulnerable Population, Model 2 Hospitals, May 2015 – June 2017

Population	Functional status: Improvement in use of mobility device (less likely to use) (pp)	Functional status: Improvement in walking without rest (pp)	Functional status: Improvement in using stairs (pp)	Care experience: Never received conflicting medical advice (pp)	Care experience: Discharged at right time[1] (pp)	Care experience: Understand care of self before going home[1] (pp)	Satisfaction with recovery[2] (pp)
All beneficiaries	-0.4	0.9	0.5	**-3.4**	-0.7	-2.1	-1.1
Dual-eligible	0.1	1.6	2.4	**-8.4**	-2.8	**-4.5**	**-7.7**
Dementia	-2.9	-1.5	3.1	**-6.0**	-3.8	-1.5	2.4
Recent institutional use	2.8	5.5	5.9	-4.0	-2.2	-2.7	-5.3

Note: The estimates in this table are the results of a cross-sectional logistic regression risk adjustment model for binary indicators. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. pp=percentage points.

[1] Measure reflects that respondents either agreed or strongly agreed with this statement.

[2] Measure reflects that respondents were either quite a bit satisfied or extremely satisfied with their recovery since leaving the hospital.

Source: Lewin analysis of beneficiary survey data for episodes that began May 2015 through June 2017 for BPCI and comparison hospitals.

Simple pneumonia and respiratory infections

Results for the claim-based analysis of vulnerable beneficiaries in the SPRI clinical episode are shown in Exhibit 26. Among all BPCI SPRI episodes, we observed a statistically significant decline in total allowed payments for BPCI beneficiaries with dementia and with recent institutional use relative to their respective comparison groups ($p<0.05$). These two vulnerable populations represent a large share of the SPRI episodes (37% and 15%, respectively). There were no statistically significant differences in emergency department use, unplanned readmissions, or mortality relative to the comparison group.

Exhibit 26: Impact of BPCI on Simple Pneumonia and Respiratory Infections Payments and Quality of Care, by Vulnerable Population, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016

Population	Total Allowed Payment Amount, IP through 90-day PDP	ED Use, 90-day PDP (pp)	Unplanned Readmission Rate, 90-day PDP (pp)	All-cause Mortality Rate, 90-day PDP (pp)
All beneficiaries	**-$689**	0.1	-0.5*	-0.2
Dual-eligible	-$254	-0.2	-0.5	1.0
Dementia	**-$1,007**	-0.1	-0.4	-0.8
Recent institutional use	**-$2,991**	-1.3	-2.2	1.7

Note: The estimates in this table are results of a difference-in-differences (DiD) model. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. IP=inpatient. PDP=post-discharge period. ED=emergency department. pp=percentage points.

* This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate. Equal trends test was conducted for total allowed payment amount and total amount paid by Medicare, IP through 90-day PDP, emergency department visits, readmission, and mortality outcomes.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

For the SPRI clinical episode, survey results do not indicate consistent findings across vulnerable populations or outcomes, though dual-eligible BPCI respondents may have had better care experiences and satisfaction with recovery than comparison respondents (Exhibit 27). In contrast to all SPRI respondents, BPCI dual-eligible respondents were more likely than the comparison group respondents to report that they never received conflicting medical advice ($p<0.05$) and were more likely to report that they were extremely or quite a bit satisfied with overall satisfaction with recovery since leaving the hospital ($p<0.10$).

Care experiences for BPCI beneficiaries with dementia may have been less favorable than comparison beneficiaries with dementia as a smaller proportion of respondents reported they had a good understanding of how to take care of themselves before going home ($p<0.10$). In addition, BPCI beneficiaries with recent institutional use were less likely to report improvements in walking without rest ($p<0.10$) and that they were extremely or quite a bit satisfied with overall satisfaction with recovery since leaving the hospital when compared to comparison group beneficiaries ($p<0.05$).

Exhibit 27: Impact of BPCI on Simple Pneumonia and Respiratory Infections Survey Measures, by Vulnerable Population, Model 2 Hospitals, May 2015 – June 2017

Population	Functional status: Improvement in use of mobility device (less likely to use) (pp)	Functional status: Improvement in walking without rest (pp)	Functional status: Improvement in using stairs (pp)	Care experience: Never received conflicting medical advice (pp)	Care experience: Discharged at right time[1] (pp)	Care experience: Understand care of self before going home[1] (pp)	Satisfaction with recovery[2] (pp)
All beneficiaries	0.4	2.2	0.7	**-3.3**	-1.6	**-2.1**	-0.2
Dual-eligible	3.5	-1.0	0.6	**7.6**	1.1	-2.6	6.2
Dementia	-0.3	-0.5	-2.1	-3.2	-1.7	-3.4	1.9
Recent institutional use	-8.1	-11.6	0.5	-0.4	-6.4	7.4	**-15.9**

Note: The estimates in this table are the results of a cross-sectional logistic regression risk adjustment model for binary indicators. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. pp=percentage points.

[1] Measure reflects that respondents either agreed or strongly agreed with this statement.

[2] Measure reflects that respondents were either quite a bit satisfied or extremely satisfied with their recovery since leaving the hospital.

Source: Lewin analysis of beneficiary survey data for episodes that began May 2015 through June 2017 for BPCI and comparison hospitals.

4. Understanding the Impact of Terminating Participation on Payments

The BPCI initiative allows participants to withdraw from a clinical episode at any time.[33] By December 2016, 53% of the Model 2 hospital EIs withdrew from at least one clinical episode.[34] Descriptive analysis suggests that hospital EIs were more likely to withdraw from clinical episodes in which they had negative NPRA, so withdrawal from BPCI is non-random. This non-random withdrawal could affect our BPCI impact estimates by disrupting the balance between the BPCI and comparison groups, as well as giving more weight to participants with a longer exposure to BPCI. It is not clear how these potential biases would affect the BPCI impact estimates. To evaluate the impact of the withdrawals on the impact estimates, we re-calculated the estimates with all episodes from participants and comparison group providers through December 2016 included in the analysis, regardless of whether the BPCI EI had withdrawn from the clinical episode prior to December 2016. We refer to these sensitivity analyses as intent-to-treat (ITT) DiD analyses.

Across all clinical episodes, 17% of matched episodes were initiated after the EIs had withdrawn from the clinical episode. The share of matched post-withdrawal episodes as a proportion of the intervention sample varied across clinical episodes, from 7% for MJRLE to 36% for gastrointestinal hemorrhage. Results by clinical episode are located in **Appendix K**.

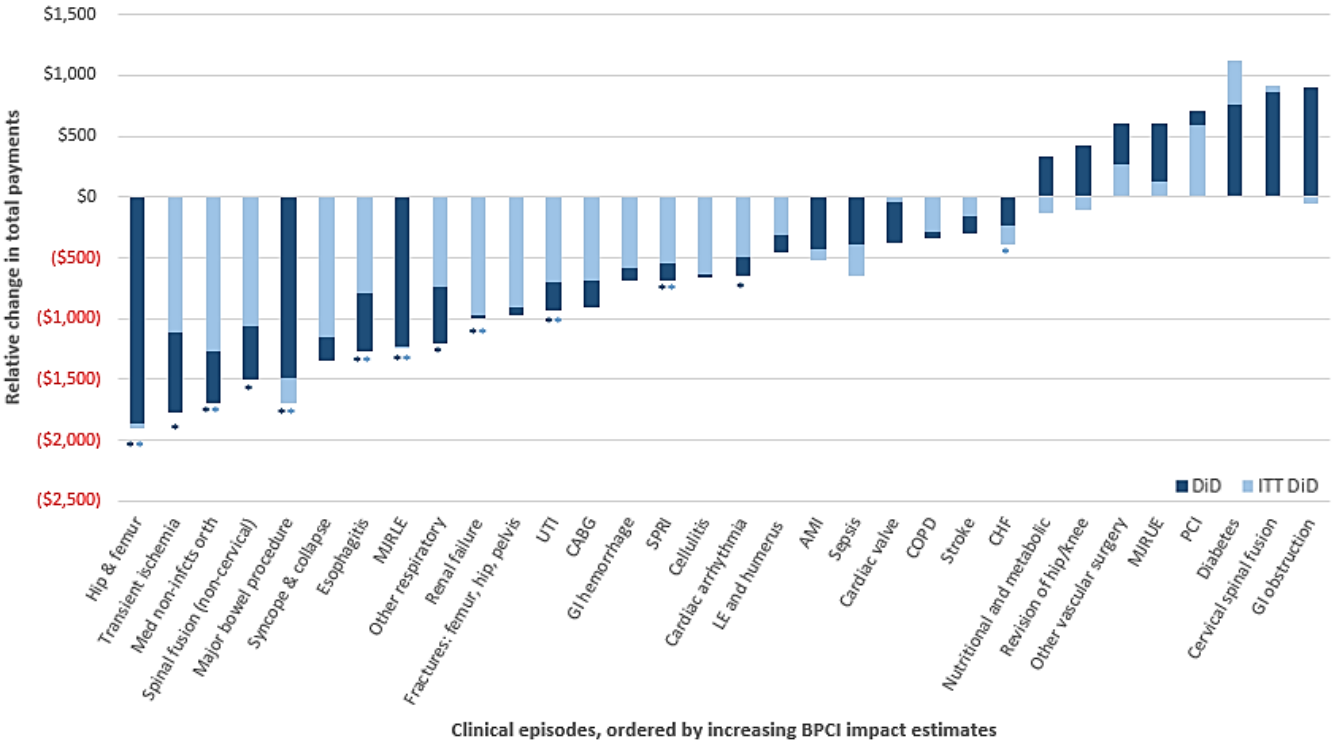
Out of 12 Model 2 hospital clinical episodes that exhibited a statistically significant relative decline in total allowed payments, eight had statistically significant ITT DiD estimates in the same direction (Exhibit 28). Four estimates remained in the same direction but were not statistically significant.[35] The ITT DiD sensitivity analysis, therefore, suggests that the BPCI impact estimates are robust to selective withdrawal for the majority of Model 2 hospital clinical episodes. For CHF, the ITT DiD estimate was statistically significant, but the BPCI impact estimate was not. This suggests that BPCI participants continued to reduce episode payments relative to the comparison group, even after they withdrew from the initiative. For additional results and methods, see **Appendix K**.

[33] We use the term withdrawal to reflect EIs that terminate their participation in BPCI or stop participating in the given clinical episode. Note that EIs that participated in a clinical episode for only one quarter are excluded from both the BPCI impact analysis and this analysis.

[34] This is limited to participants that met the minimum number of patient episodes to be included in our DiD and intent-to-treat (ITT) DiD analysis.

[35] Furthermore, nine of the 12 clinical episodes that exhibited significant relative declines in the original impact estimate exhibited smaller relative declines in the ITT DiD estimate. This suggests that in certain cases the BPCI effect may have diminished after withdrawal.

Exhibit 28: BPCI Impact Estimates and ITT DiD Estimates for Total Medicare Allowed Payments during the Inpatient Stay and 90-days Post discharge, by Clinical Episode, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q4 2016



Note: The estimates in this exhibit are the results of the BPCI impact difference-in-differences (DiD) model (dark blue) and intent-to-treat (ITT) DiD model (light blue). The results are ranked by increasing magnitude of the BPCI impact estimates. BPCI impact estimates that are significant at the 10% significance level are indicated by a dark blue asterisk; ITT DiD estimates that are significant at the 10% significance level are indicated by a light blue asterisk. The Medicare payment outcome is standardized to remove the effect of geographic and other adjustments. MJRLE=major joint replacement of the lower extremity. UTI=urinary tract infection. CABG=coronary artery bypass graph. GI=gastrointestinal. SPRI=simple pneumonia and respiratory infections. LE=lower extremity. AMI=acute myocardial infarction. COPD=chronic obstructive pulmonary disease. CHF=congestive heart failure. MJRUE=major joint replacement of the upper extremity. PCI=percutaneous coronary intervention.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

5. Market Dynamics

This section presents the analysis of market share and provider referral patterns for Model 2 hospitals.

a. Key Findings

- Overall, there was a small decline in BPCI market share from the baseline period to the intervention period.
- There was no change in the number of SNF or HHA providers to which Model 2 hospital EIs discharged their patients.
- There was a moderate increase in the share of Model 2 hospital patients who were discharged to SNF providers with high star-ratings in the intervention period and a negligible change in the share of patients who were discharged to HHA providers with

high star-ratings, though results varied considerably by clinical episode and BPCI participant cohorts.

b. Methods

We conducted a descriptive analysis to understand whether BPCI Model 2 EIs captured a greater share of the episodes in their markets for the clinical episodes in which they participated. We used the Medicare Core-Based Statistical Areas (MCBSAs) to define BPCI EIs' markets, excluding two markets that were likely too large to accurately define the local health care market (Chicago-Naperville-Arlington Heights, IL and Cincinnati-Middletown, OH-KY-IN). We also examined the number of PAC providers to which Model 2 hospital EIs discharged their patients, as well as the share of patients discharged from Model 2 hospital EIs who subsequently received PAC services at PAC providers with high quality ratings (based on the CMS star ratings of SNFs and HHAs).

Market shares and referral patterns may be considerably different for different clinical episodes and types of EIs. In addition, BPCI allowed rolling enrollment and withdrawal of participants. Recognizing these technical challenges, we stratified the market dynamics analysis by BPCI clinical episode and EI cohort. An EI cohort includes all BPCI EIs that entered Phase 2 in a clinical episode in the same quarter. To detect meaningful changes in market measures over time, we focused on Model 2 hospital clinical episodes with 30 or more EIs in any cohort of interest. For each episode and EI cohort, we calculated the market measure rate in the baseline period (Q4 2011 through Q3 2012), as well as the change from the baseline period to each six-month interval in the intervention period.

In addition to analyses stratified by clinical episodes, cohorts, and six-month intervals, we also calculated aggregate measures by combining clinical episodes, cohorts, and six-month intervals to discern the overall trends for Model 2 hospital EIs – a single rate for each measure. To calculate the aggregate measures, we combined the latest two six-month intervals in the intervention period (referred to as the "last BPCI year in the analysis" hereafter) and compared them to the baseline period.

We used the median of each market dynamics measure to discern changes over time, instead of the mean, because the median is less sensitive to outliers. The full distribution (including mean, median, 25th percentile, and 75th percentile) of the market dynamics measures is reported in **Appendix L**.

c. Results

The market share for BPCI Model 2 hospital EIs declined slightly over time. Across the nine clinical episode-cohort combinations, Model 2 hospital EIs had a median market share of 8.5% in the one-year baseline period (Q4 2011 through Q3 2012). The median change in market share across these Model 2 hospital EIs was -0.1 pp between the baseline period and the last BPCI year in this analysis (Exhibit 29). Median changes in the market share of Model 2 hospital EIs varied by clinical episode and cohort, as well as by cohort within the same clinical episode, e.g., MJRLE, though the changes were relatively small for all results (see **Appendix L** for results by clinical episode and cohort).

Overall, Model 2 hospital EIs discharged 75% of their patients to a median of 4.5 SNF providers and 2.5 HHA providers in the baseline period. There was no change in the number of SNF or HHA providers used by patients discharged from BPCI-participating hospitals between the baseline period and the last BPCI year in this analysis (Exhibit 29). However, the results varied by clinical episode (see **Appendix L** for results).

Over time, there was a moderate increase in the use of highly rated SNF providers and a negligible increase in the use of highly rated HHA providers following BPCI hospital discharges. Overall, Model 2 hospital EIs discharged 25.0% of their patients to a SNF with a high star rating in the baseline period and 26.1% of their patients to a HHA with a high star rating. Across all Model 2 hospital EIs during the last BPCI year in this analysis, the median increase in the share of beneficiaries discharged to a SNF or HHA with a high star rating was 3.9 pp and 0.4 pp, respectively (Exhibit 29). At the clinical episode and EI cohort level, there was a consistently larger share of BPCI hospital discharges to highly rated SNFs among seven of the nine combinations. There was no consistent change in the percent of Model 2 hospital discharges to high star rating HHAs, with one exception: a decline for hip and femur procedures except major joint that entered Phase 2 in Q4 2015, though small sample sizes makes it difficult to interpret the changes over time (see **Appendix L** for results).

Exhibit 29: Median Baseline Rate and Median Changes from Baseline for Market Dynamics Measures, Model 2 Hospitals EIs, Q4 2011 – Q4 2016

Measures	Median Baseline Rate	Median Changes from Baseline
EI Market Share (%)	8.5	-0.1
Number of PAC Providers Used: SNFs	4.5	0.0
Number of PAC Providers Used: HHAs	2.5	0.0
Percent of PAC Admissions to High Star-Rating SNFs (%)	25.0	3.9
Percent of PAC Admissions to High Star-Rating HHAs (%)	26.1	0.4

Note: This table shows the measure rate in the baseline period, and the changes from the baseline period to the last BPCI year in the analysis. All selected Model 2 hospital clinical episodes and cohorts are combined. EI = episode initiator. SNF = skilled nursing facility. HHA = home health agency. PAC = post-acute care.

Source: Lewin analysis of Medicare claims for discharges that began Q4 2011 through Q4 2016.

C. Impact of BPCI among Participating PGPs

1. Payment, Utilization, Quality, and Patient Mix

This section presents the BPCI impact estimates on payments, utilization, and quality, as well as characteristics of the mix of patients in PGP-initiated episodes for the first 13 quarters of the initiative (Q4 2013 through Q4 2016).

a. Key Findings

- There was a relative decline in total Medicare payments during the inpatient stay plus 90 days post discharge for 16 (7 were statistically significant) of the 21 Model 2 PGP clinical episodes for which we had sufficient sample size during the first 13 quarters of the BPCI initiative; 7 were statistically significant.

- Across many of the clinical episodes, PAC payments, especially SNF payments, declined. Smaller shares of episodes were discharged to institutional PAC settings, and there were fewer average SNF days for SNF users. HHA and readmission payments also declined for a number of clinical episodes.
- In general, rates of ED use, readmissions, and mortality did not change under BPCI.
- The majority of PGP clinical episodes had no indication that the BPCI patient mix became more or less resource intensive than the comparison group during the intervention period.

b. Methods

Similar to the methods described in Section II.B.1 above, the analysis uses a DiD design to estimate the differential change in payment, utilization, and quality outcomes between the baseline and an intervention period for beneficiaries who received services from BPCI providers relative to beneficiaries who received services from a comparison group of non-BPCI providers. This approach controls for health care service use before the hospitalization, beneficiary, market, and provider differences between BPCI and comparison episodes; eliminates biases from time invariant differences between the BPCI and comparison episodes; and controls for common trends in the BPCI and comparison population. Because we did not have reliable data on physician affiliation to create a non-BPCI PGP comparison group, we constructed a comparison group of hospitals similar to the hospitals associated with BPCI PGPs to minimize differences in the distributions of characteristics between the hospitals associated with BPCI PGPs and comparison hospitals. Each BPCI PGP episode was randomly matched by quarter and MS-DRG to an episode from a comparison hospital.

We constructed comparison groups for 21 Model 2 PGP clinical episodes that had sufficient sample size for meaningful analysis. We considered a clinical episode to have sufficient sample size if there were 20 EIs with a total of 1,000 episodes initiated during the intervention, however, this minimum sample size was not based on a formal power calculation. These 21 clinical episodes contained 93% of all episodes initiated by Model 2 PGP EIs during the first 13 quarters of the initiative. See **Appendix D** for additional details on the methodology.

c. Results

Sample characteristics

We describe the BPCI Model 2 PGP analytical sample for the clinical episodes for which we had sufficient sample size in Exhibit 30. The number of BPCI PGP EIs included in the sample ranged from 20 to 113, and the number of episodes ranged from 1,563 to 64,392, depending on the clinical episode, from the beginning of BPCI in Q4 2013 through Q4 2016. Providers could join BPCI over an extended period and could terminate participation in a clinical episode or withdraw from the initiative completely at any time. As a result, providers in this analysis participated for an average of five quarters from Q4 2013 to Q4 2016. During this period, 71 of 272 Model 2 PGP EIs (26%) withdrew entirely from the initiative. Among the 190 PGP EIs participating in any of the 21 clinical episodes analyzed in this report, approximately 56% stopped participating in at least one clinical episode by Q4 2016. The BPCI impact estimates include episodes from all PGP EIs during their participation period. The contribution of episodes from PGP EIs that

stopped participating by Q4 2016 varies by clinical episode. For four clinical episodes, PGPs that stopped participating in the clinical episode had contributed 50% or more of the episodes during the intervention period. For an additional seven clinical episodes, the PGP EIs that stopped participating contributed over 40% of the episodes. The impact of withdrawal on BPCI is discussed in detail in Section II.C.3 below.

Exhibit 30: Characteristics of the Matched BPCI Providers Included in the BPCI Impact Estimates, Model 2 PGP, Q4 2013 – Q4 2016

Clinical Episode	BPCI PGPs (#)	Matched Intervention Episodes (#)	Average Length of Participation (Quarters)	EIs that Terminated Participation in the Clinical Episode (#)	Episodes from EIs that Terminated (%)
Acute myocardial infarction	51	3,772	5	19	46.9
Cardiac arrhythmia	35	2,978	6	11	27.9
Cellulitis	52	2,859	5	22	40.7
Congestive heart failure	46	9,584	5	16	39.0
COPD, bronchitis, asthma	59	9,200	5	24	47.3
Esophagitis, gastroenteritis & other digestive disorders	43	3,723	5	17	40.1
Gastrointestinal hemorrhage	36	2,461	5	14	41.5
Gastrointestinal obstruction	37	1,563	5	15	29.8
Hip & femur procedures except major joint	64	6,328	6	25	39.2
Major joint replacement of the lower extremity	113	64,392	6	29	14.5
Major joint replacement of the upper extremity	32	2,476	5	11	29.4
Medical non-infectious orthopedic	42	3,437	5	20	54.8
Nutritional and metabolic disorders	43	2,674	6	14	38.7
Other respiratory	49	3,742	5	19	25.4
Percutaneous coronary intervention	29	2,722	6	10	51.5
Renal failure	43	5,899	5	19	53.6
Sepsis	63	20,802	5	27	41.9
Simple pneumonia & respiratory infections	61	10,117	5	28	46.1
Spinal fusion (non-cervical)	20	2,391	5	8	18.9
Stroke	38	3,485	6	13	32.0
Urinary tract infection	51	6,422	6	21	61.0

Note: 37% of the BPCI PGPs were not included in the sample. This exhibit is limited to the BPCI providers used to calculate the DiD results in the remainder of this section. See **Appendix D** for information on the methods used to determine the sample. DiD=difference-in-differences. PGP=physician group practice. EI=episode initiator. COPD=chronic obstructive pulmonary disease.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2013 through Q4 2016 for BPCI providers.

Has patient mix changed under BPCI?

Because target prices are not risk adjusted, episode payments for BPCI participants could decline relative to the target price if the mix of patients during the intervention period was less resource intensive than the mix of patients during the baseline. Similarly, episode payments could increase relative to the target price if patient mix was more resource intensive in the intervention than in the baseline. To assess whether BPCI participant patient mix changed during the intervention, we examined claim-based patient characteristics that are associated with patient resource use. Exhibit 31 shows estimates of the change between the baseline and intervention period for BPCI patients relative to the change in the comparison group for demographic characteristics, count of HCCs, and the utilization of care in the six months prior to the anchor hospitalization. For each of the measures in Exhibit 31, a negative value indicates a decline in the resource intensity of the BPCI patients during the intervention from the baseline period relative to the comparison group. Similarly, a positive value suggests a relative increase in patient resource intensity. (Please note: the impact analysis on payment, utilization, and quality presented below controlled for changes in these patient characteristics.)

We categorized Model 2 PGP strata (including MJRLE episodes stratified into fractures and non-fractures), into three broad groups: decline in patient resource intensity, increase in patient resource intensity, and no change.[36] Our categorization was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates. See **Appendix D** for additional details on the categorization.

Of the 22 Model 2 strata, the majority (19) did not have a consistent pattern of changes across measures. The non-fracture MJRLE stratum, however, had indications that BPCI patients were less resource intensive in the intervention period relative to the change for the comparison group. MJRLE not due to fracture is a planned, elective procedure, so participants could choose who they treated (see **Appendix F** for more details on which clinical episodes are planned, elective procedures). The resource intensity of the medical non-infectious orthopedic and urinary tract infection strata in the BPCI-participating PGPs may have increased relative to patient mix in the comparison group.

[36] The "no change" category includes strata that do not exhibit a consistent pattern toward a decline or an increase in patient resource intensity. This could be because they have indications of both decreases and increases in patient resource intensity or no indications of changes in either direction.

Exhibit 31: Relative Changes in Patient Resource Intensity, by Clinical Episode Strata, Baseline to Intervention, Model 2 PGP, Q4 2013 – Q4 2016

Change in Patient Mix	Clinical Episode Strata	Number of Episodes Q4 2013- Q4 2016	Age: 80+ years	Medicaid Eligibility	Disabled, no ESRD	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
Less resource intensive	Major joint replacement of the lower extremity - Non-fractures	56,628	-0.2	-0.1	-1.0	-0.02	-0.6	-0.9	-0.1	-0.4
More resource intensive	Medical non-infectious orthopedic	3,437	-4.3	3.0	4.1	0.13	2.9	6.3	0.2	0.9
	Urinary tract infection	6,422	-0.6	2.1	0.6	0.15	3.9	0.5	2.4	0.0
No consistent pattern	Acute myocardial infarction	3,772	-5.6	-2.9	3.5	0.03	2.0	-0.6	-2.4	-0.9
	Cardiac arrhythmia	2,978	0.3	-0.5	-0.7	0.04	2.2	-1.2	-0.2	1.3
	Cellulitis	2,859	1.7	2.4	1.6	0.04	2.0	4.4	-0.3	-0.3
	COPD, bronchitis, asthma	9,200	-0.1	0.8	-0.6	0.17	2.3	-4.0	1.2	0.8
	Congestive heart failure	9,584	1.3	-0.4	0.2	0.05	0.6	-1.4	-0.2	0.6
	Esophagitis, gastroenteritis and other digestive disorders	3,723	0.8	0.4	0.9	0.06	0.6	1.1	-0.3	0.5
	Gastrointestinal hemorrhage	2,461	2.2	-2.2	0.3	-0.01	0.6	-0.2	5.1	1.6
	Gastrointestinal obstruction	1,563	-1.4	-0.2	-0.1	0.03	0.1	-2.2	2.9	-0.8
	Hip & femur procedures except major joint	6,328	-0.9	0.8	0.1	0.04	1.4	-1.0	1.8	-0.4
	Major joint replacement of the lower extremity - Fractures	7,764	0.5	-0.1	-0.2	0.04	0.7	0.6	2.0	0.0
	Major joint replacement of the upper extremity	2,476	-4.6	-1.6	1.7	-0.08	-1.0	0.7	-1.0	-1.4
	Nutritional and metabolic disorders	2,674	-1.7	-0.4	-0.1	-0.05	0.4	-1.2	2.9	-1.8
	Other respiratory	3,742	2.7	-2.3	-3.4	0.05	2.7	0.3	-2.3	1.0
	Percutaneous coronary intervention	2,722	3.1	-0.4	-2.5	-0.09	1.7	-2.4	-0.9	0.2
	Renal failure	5,899	-1.0	1.7	2.1	0.11	1.2	-0.4	-1.5	-3.3
	Sepsis	20,802	-0.2	-0.4	-1.5	0.11	0.1	-0.7	0.5	0.2

Change in Patient Mix	Clinical Episode Strata	Number of Episodes Q4 2013- Q4 2016	Age: 80+ years	Medicaid Eligibility	Disabled, no ESRD	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
No consistent pattern (cont'd)	Simple Pneumonia and respiratory infections	10,117	0.8	-1.7	0.2	-0.04	0.2	-2.1	-0.2	-0.2
	Spinal fusion (non-cervical)	2,391	1.2	0.5	-1.1	0.04	-0.2	-3.6	-0.8	-0.6
	Stroke	3,485	1.0	-1.6	-1.7	0.09	-0.1	1.1	3.7	-1.2

Note: Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Categorization of resource intensity was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates. See Appendix D for additional details of the categorization. ESRD=end-stage renal disease. HCC=hierarchical conditions categories. COPD=chronic obstructive pulmonary disease.

* These characteristics measure utilization of care in the six months prior to the anchor hospitalization. Count of HCCs is based on the six months prior to the anchor hospitalization.

Source: Lewin analysis of Medicare claims and enrollment for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q4 2016 (intervention period) for BPCI EIs and the matched comparison providers.

How have the average standardized payments changed under BPCI?

Since the implementation of BPCI in Q4 2013 through Q4 2016, the total allowed payment amount for the inpatient stay plus the 90-day PDP declined from the baseline to intervention period for the majority of BPCI PGP-initiated clinical episodes relative to comparison episodes. Reduced PAC payments, particularly for SNF, were the major contributor to these declines, with readmission payments also contributing to reduced payments in some clinical episodes (Exhibits 32 and 33). Detailed results of the BPCI impact estimates by clinical episode are located in **Appendix M**.

The total allowed payment amount for the inpatient stay plus 90-day PDP decreased in 16 of 21 clinical episodes; the decline for seven clinical episodes was statistically significant ($p<0.10$). Across these seven clinical episodes, BPCI PGPs reduced total allowed payments by 6.9% on average relative to their payments absent BPCI (Exhibit 34).[37] These declines were primary due to reductions in PAC during the 90-day PDP. The SNF payments decreased for 15 of the 21 clinical episodes, and the decline was statistically significant in five ($p<0.10$). Statistically significant reductions in SNF payments were associated with statistically significant reductions in total payments for three of the five clinical episodes (Exhibit 33). The IRF payments decreased relative to the change in the comparison group for six of the 14 clinical episodes with sufficient sample size to examine; the decline in two clinical episodes was statistically significant ($p<0.10$). The HHA payments decreased relative to the change for the comparison group in 11 of the 21 clinical episodes; the decrease was statistically significant in four ($p<0.10$). One clinical episode had a statistically significant relative increase in HHA payments ($p<0.05$).

There was a statistically significant increase in the total allowed payment amount for the inpatient stay plus 90-day PDP for cardiac arrhythmia. It was also the only clinical episode with a statistically significant increase in the SNF standardized allowed amount.

For the five clinical episodes that had the highest proportion of total baseline payments for PAC services (hip and femur procedures except major joint, MJRLE, medical non-infectious orthopedic, stroke, and urinary tract infection), total allowed payments for the inpatient stay plus the 90-day PDP declined relative to the change in comparison episodes in all five clinical episodes; the decline was statistically significant in three ($p<0.10$). The SNF payments declined in all five clinical episodes, and the decline in one clinical episode was statistically significant. The IRF payments declined for four of the five clinical episodes, and the decline was statistically significant in two clinical episodes ($p<0.10$). MJRLE had a statistically significant relative decline in SNF, HHA, and IRF payments ($p<0.10$). See Exhibit 32 below or **Appendix F** for additional details on which clinical episodes have a larger proportion total baseline payments for PAC services.

[37] BPCI provider payments absent BPCI are calculated as what their payments would have been if they had experienced the same change in payments from the baseline to intervention period as the comparison group.

Exhibit 32: Impact of BPCI on Payment Outcomes, by Clinical Episode, Model 2 PGP, Baseline to Intervention, Q4 2013 – Q4 2016

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Total Allowed Amount Included in Bundle Definition[1]	Total Amount Paid by Medicare, IP through 90-day PDP[2]	Total Allowed Payment Amount, IP through 90-day PDP	SNF Standardized Allowed Amount, 90-day PDP[3]	HHA Standardized Allowed Amount, 90-day PDP[3]	IRF Standardized Allowed Amount, 90-day PDP[3]	Readmissions Standardized Allowed Amount, 90-day PDP[3]	Part B Standardized Allowed Amount, 90-day PDP	Total Part A & B 30-day Post-bundle
Acute myocardial infarction	3,772	-$51	$4	-$159	-$406	-$118	$193	$202	-$215	-$308
Cardiac arrhythmia	2,978	$1,086	$922	$1,148	**$567**	-$19	$202	$162	$48	-$145
Cellulitis	2,859	$149	-$236	-$262	-$496	**-$179**		$91	-$28	-$174
Congestive heart failure	9,584	-$408	-$116	-$200	-$278	$72	$79	-$68	-$44	-$396
COPD, bronchitis, asthma	9,200	$34	$21	-$73	-$312	$13	$102	$206	-$119	-$38
Esophagitis, gastroenteritis & other digestive disorders	3,723	$343	$329*	$269*	$22	$37		$60	-$86	**$437**
Gastrointestinal hemorrhage	2,461	-$1,119	**-$1,397**	**-$1,531**	-$277	$64		-$673	-$269	$53
Gastrointestinal obstruction	1,563	**-$2,134**	**-$1,750**	**-$2,024**	$6	-$15		**-$1,020**	**-$599**	-$394
Hip & femur procedures except major joint[4]	6,328	**-$1,176**	**-$1,082**	**-$1,241***	-$321	$29	-$646	-$37	-$66	-$273
Major joint replacement of the lower extremity[4]	64,392	**-$1,960**	**-$1,848**	**-$1,958**	**-$1,084**	**-$362**	**-$616**	-$71	**-$109**	$1
Major joint replacement of the upper extremity	2,476	**-$1,857**	**-$1,763**	**-$2,065**	**-$1,681**	-$186		-$83	-$31	-$182
Medical non-infectious orthopedic[4]	3,437	-$1,353	-$1,184	-$1,472	-$486	-$92	-$319	$13	-$180	**-$540**

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Total Allowed Amount Included in Bundle Definition[1]	Total Amount Paid by Medicare, IP through 90-day PDP[2]	Total Allowed Payment Amount, IP through 90-day PDP	SNF Standardized Allowed Amount, 90-day PDP[3]	HHA Standardized Allowed Amount, 90-day PDP[3]	IRF Standardized Allowed Amount, 90-day PDP[3]	Readmissions Standardized Allowed Amount, 90-day PDP[3]	Part B Standardized Allowed Amount, 90-day PDP	Total Part A & B 30-day Post-bundle
Nutritional and metabolic disorders	2,674	-$914	-$661	-$879	-$863	-$34		-$447	-$158	-$85
Other respiratory	3,742	-$515	-$140	-$251	-$101	-$30	$85	-$434	-$121	-$225
Percutaneous coronary intervention	2,722	$60	-$331	-$438	$175	-$48		-$163	-$167	-$127
Renal failure	5,899	$578	$490	$539	$18	**$125**	$92	$13	$48	$93
Sepsis	20,802	$28	$99	$54	-$102	$13	-$56	-$88	$50	-$81
Simple pneumonia & respiratory infections	10,117	-$259	$157	$105	$95	$16	$22	$26	-$105	-$52
Spinal fusion (non-cervical)	2,391	**-$2,068**	**-$2,313**	**-$2,507**	**-$821**	**-$266**	-$440	-$370	-$209	-$167
Stroke[4]	3,485	-$1,018	-$643	-$689	-$133	$99	-$470	-$77	-$173	**$606**
Urinary tract infection[4]	6,422	-$563	-$510	-$683	-$534	$74	$43	-$149	-$148	-$103

Note: The estimates in this table are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented due to insufficient sample size. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. PGP=physician group practice. IP=inpatient. PDP=post-discharge period. SNF=skilled nursing facility. HHA=home health agency. IRF=Inpatient Rehabilitation Facility. COPD = chronic obstructive pulmonary disease.

[1] The total amount included in the bundle definition is based on only the 90-day episodes.

[2] Total amount paid does not include beneficiary out-of-pocket expenses.

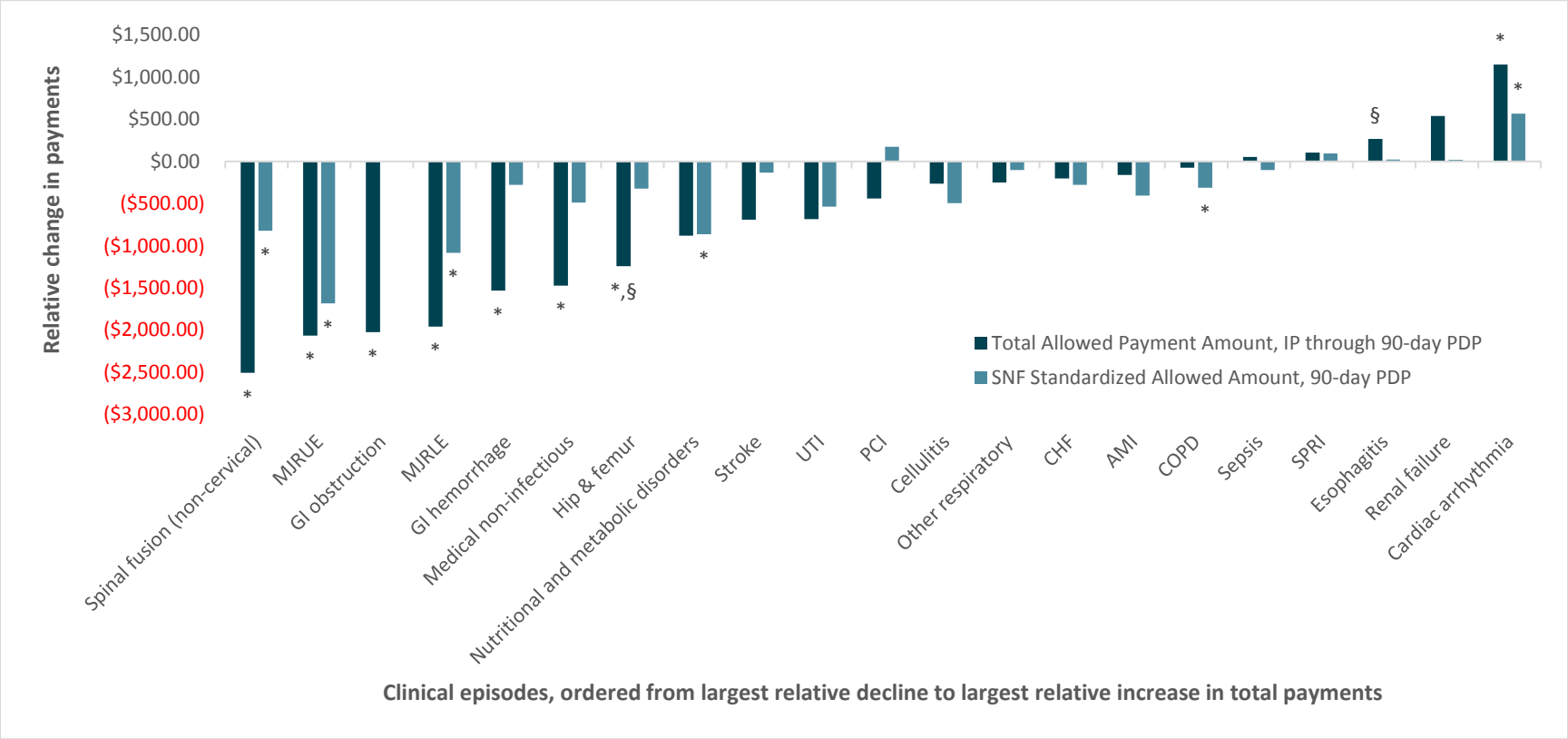
[3] These payment measures are not conditional upon use of the service.

[4] This clinical episode is one of the five clinical episodes with the highest proportion of total baseline episode payments due to PAC payments.

*This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate. Equal trends test was conducted for total allowed payment amount and total amount paid by Medicare IP through 90-day PDP, emergency department visits, readmission, and mortality outcomes.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 33: Impact of BPCI on SNF Payments and Total Medicare Allowed Payments, by Clinical Episode, Model 2 PGP, Baseline to Intervention, Q4 2013 – Q4 2016



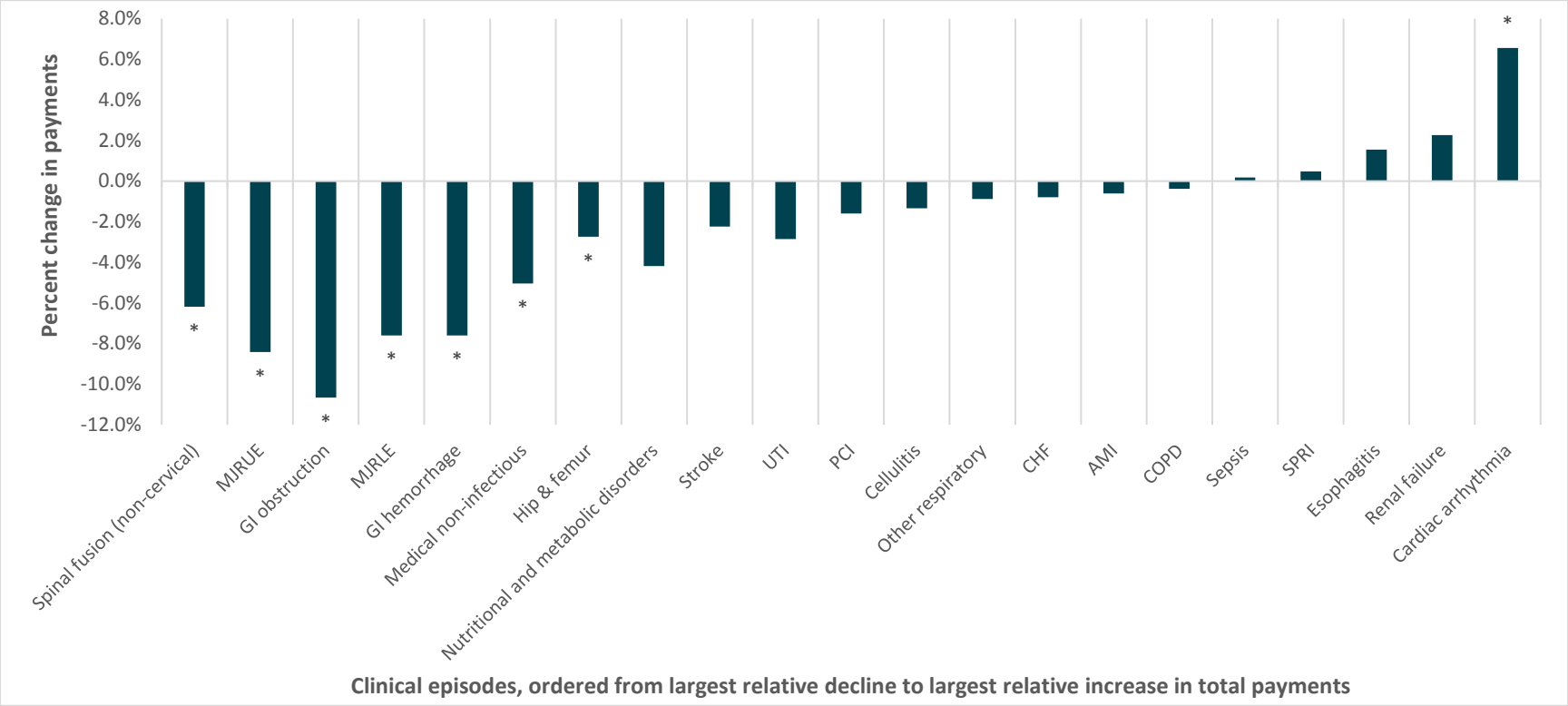
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and ordered starting with the largest relative decline in total allowed payment amount, IP-through 90-day PDP. These payment measures are not conditional upon the use of the service. SNF=skilled nursing facility. PGP=physician group practice. MJRUE=major joint replacement of the upper extremity. GI=gastrointestinal. MJRLE=major joint replacement of the lower extremity. UTI=urinary tract infection. PCI=percutaneous coronary intervention. CHF=congestive heart failure. AMI=acute myocardial infarction. COPD=chronic obstructive pulmonary disease. SPRI=simple pneumonia and respiratory infections.

*Indicates DiD estimates are statistically significant at the 10% level.

§ Data from the baseline period shows BPCI and matched comparison providers were not on parallel trends for this outcome, which is required for an unbiased estimate.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 34: Percent Change in BPCI Episode Allowed Payments from What Payments Would have Been Absent BPCI, by Clinical Episode, Model 2 PGP, Q4 2013 – Q4 2016



Note: The payments in this exhibit are the risk-adjusted standardized allowed amounts for the inpatient stay plus 90-day PDP and ordered starting with the largest relative decline. Episode payments absent BPCI, or the counterfactual, is the BPCI baseline payment amount plus the change in episode payment amount for the comparison group. The counterfactual can be expressed as: BPCI before + (Comparison after – Comparison before). The percent change can then be expressed as: (BPCI after – Counterfactual) / (Counterfactual). Results are sorted by the total Medicare allowed payment amount DiD estimate. MJRUE=major joint replacement of the upper extremity. GI=gastrointestinal. MJRLE=major joint replacement of the lower extremity. UTI=urinary tract infection. PCI=percutaneous coronary intervention. CHF=congestive heart failure. AMI=acute myocardial infarction. COPD=chronic obstructive pulmonary disease. SPRI=simple pneumonia and respiratory infections.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

How have the services changed under BPCI?

Across all 21 clinical episodes, the inpatient hospital length of stay (LOS) increased in 13 clinical episodes relative to the comparison episodes, but the increase was statistically significant for two clinical episodes ($p<0.10$). There was a relative decline in LOS for the other eight clinical episodes, and the decline was statistically significant in three ($p<0.10$).

The changes in PAC service use were consistent with the changes in payments (Exhibit 35). For 14 of the 21 clinical episodes, there was a relative decline in the proportion of patients discharged to any PAC (SNF, IRF, LTCH, or HHA), and this decline was statistically significant for three clinical episodes ($p<0.10$). Of the episodes discharged to PAC, there was a relative decline in the proportion of BPCI episodes discharged to institutional PAC (SNF, IRF or LTCH) for 10 of the 21 clinical episodes, and the decline was statistically significant for two ($p<0.05$) (Exhibit 36). For BPCI episodes that used SNF care, the number of SNF days declined in 16 clinical episodes relative to comparison episodes, and the decline was statistically significant in five ($p<0.10$) (Exhibit 37). The number of home health visits among BPCI episodes that had at least one visit increased from the baseline to the intervention period relative to the change in the comparison group in 13 clinical episodes and this difference was statistically significant in two ($p<0.05$).

There was a clear pattern with changes in PAC use in the five clinical episodes with the highest proportion of PAC payments in the baseline period. In all of these clinical episodes, the number of SNF days for SNF users declined relative to comparison episodes, and the decline was statistically significant for two clinical episodes ($p<0.05$). The number of HHA visits for HHA users increased for four clinical episodes and was statistically significant for one clinical episode ($p<0.05$).

Exhibit 35: Impact of BPCI on Utilization Outcomes, by Clinical Episode, Model 2 PGP, Baseline to Intervention, Q4 2013 – Q4 2016

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Anchor Hospital Stay LOS	Percent Discharged to PAC	Percent Discharged to an Institution out of Those who Received any PAC[1]	Number of SNF Days, 90-day PDP[2]	Number of HHA Visits, 90-day PDP[3]
Acute myocardial infarction	3,772	**0.29**	-0.8	2.9	-2.8	-0.9
Cardiac arrhythmia	2,978	0.01	2.3	**6.3**	3.1	0.2
Cellulitis	2,859	0.13	-3.3	1.8	-2.6	-1.1
Congestive heart failure	9,584	-0.10	2.5	-1.5	0.0	**0.9**
COPD, bronchitis, asthma	9,200	**0.16**	0.7	0.8	**-2.5**	0.2
Esophagitis, gastroenteritis & other digestive disorders	3,723	0.11	0.9	1.5	-1.2	-0.4
Gastrointestinal hemorrhage	2,461	0.11	-0.6	-6.5	-0.2	0.1
Gastrointestinal obstruction	1,563	-0.23	0.6	2.1	0.6	0.4
Hip & femur procedures except major joint[4]	6,328	0.07	-0.5	0.2	**-2.2**	0.4
Major joint replacement of the lower extremity[4 5]	64,392	**-0.09**	**-8.1**	**-3.9**	**-4.0**	**-1.2**

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Anchor Hospital Stay LOS	Percent Discharged to PAC	Percent Discharged to an Institution out of Those who Received any PAC[1]	Number of SNF Days, 90-day PDP[2]	Number of HHA Visits, 90-day PDP[3]
Major joint replacement of the upper extremity[5]	2,476	-0.06	-5.1	**-9.5**	**-11.1**	-1.0
Medical non-infectious orthopedic[4]	3,437	-0.06	-3.1	-1.3	-1.2	0.0
Nutritional and metabolic disorders	2,674	0.11	-0.9	-3.8	-1.6	0.3
Other respiratory	3,742	0.03	-0.1	3.2	-1.9	0.9
Percutaneous coronary intervention[5]	2,722	-0.19	-0.8	6.8	-2.2	-0.6
Renal failure	5,899	0.01	-0.3	-2.8	1.5	0.6
Sepsis	20,802	0.14	-0.3	-1.4	0.1	0.3
Simple pneumonia & respiratory infections	10,117	0.04	1.1	1.6	0.0	0.4
Spinal fusion (non-cervical)[5]	2,391	-0.01	**-6.8**	-0.7	-3.3	-1.0
Stroke[4]	3,485	0.08	-0.5	0.4	-0.3	0.8
Urinary tract infection[4]	6,422	-0.02	0.2	-2.1	-1.3	**1.8**

Note: The estimates in this table are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. PGP=physician group practice. PAC=post-acute care. LOS=length of stay. SNF=skilled nursing facility. PDP=post-discharge period. HHA=home health agency. COPD= chronic obstructive pulmonary disease.

[1] Institutional PAC includes SNF, IRF, and LTCH. Non-institutional PAC includes HHA.

[2] Beneficiaries must have spent a minimum of one day in a SNF setting during the 90-day PDP.

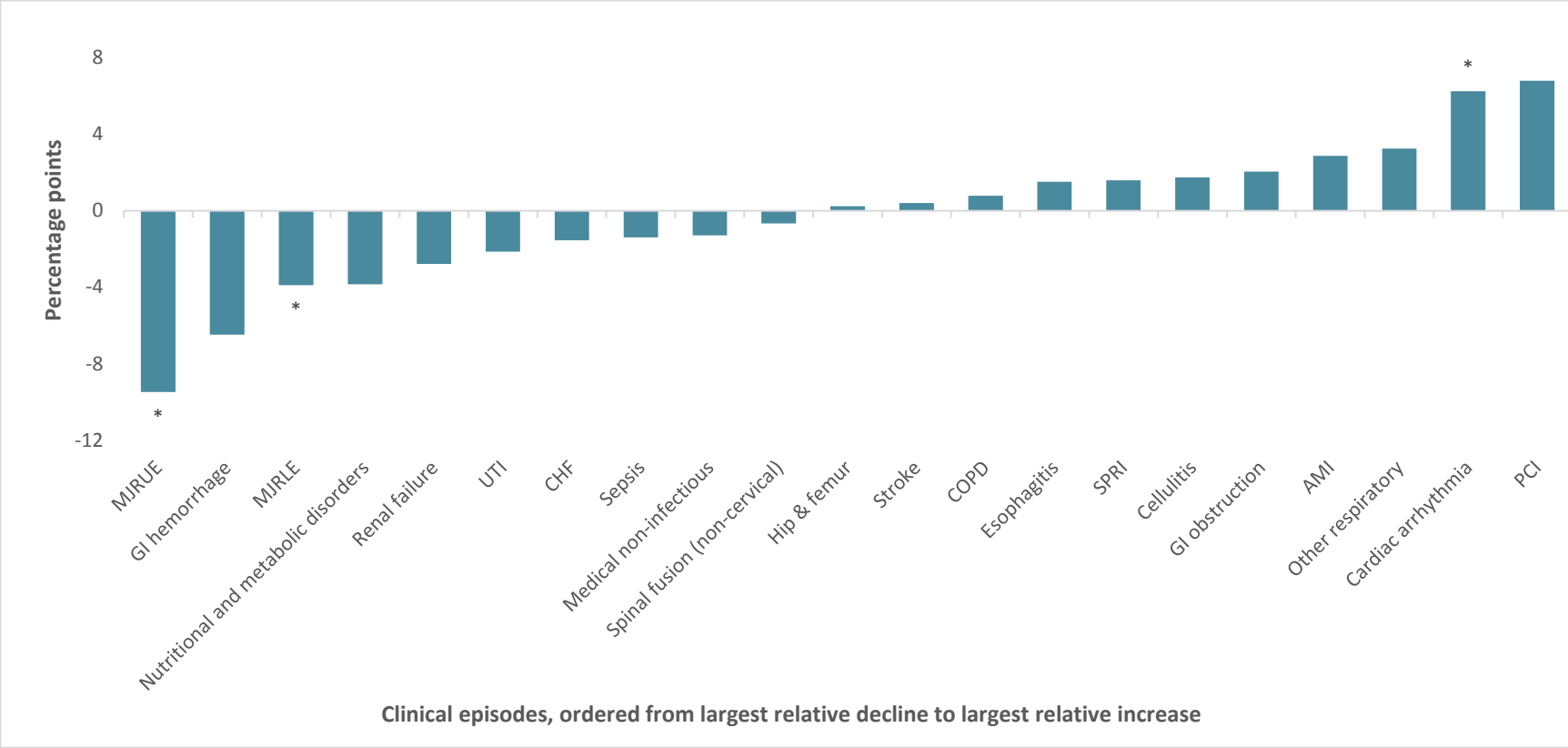
[3] Beneficiaries must have had a minimum of one HHA visit during the 90-day PDP.

[4] This clinical episode is one of the 10 with the highest proportion of total baseline episode payments due to PAC payments.

[5] This clinical episode is one of the nine with the highest proportion of total baseline episode payments due to the anchor inpatient stay (>40%).

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 36: Impact of BPCI on the Percent of Beneficiaries Discharged to Institutional PAC out of those who Received Any PAC, by Clinical Episode, Model 2 PGP, Baseline to Intervention, Q4 2013 – Q4 2016

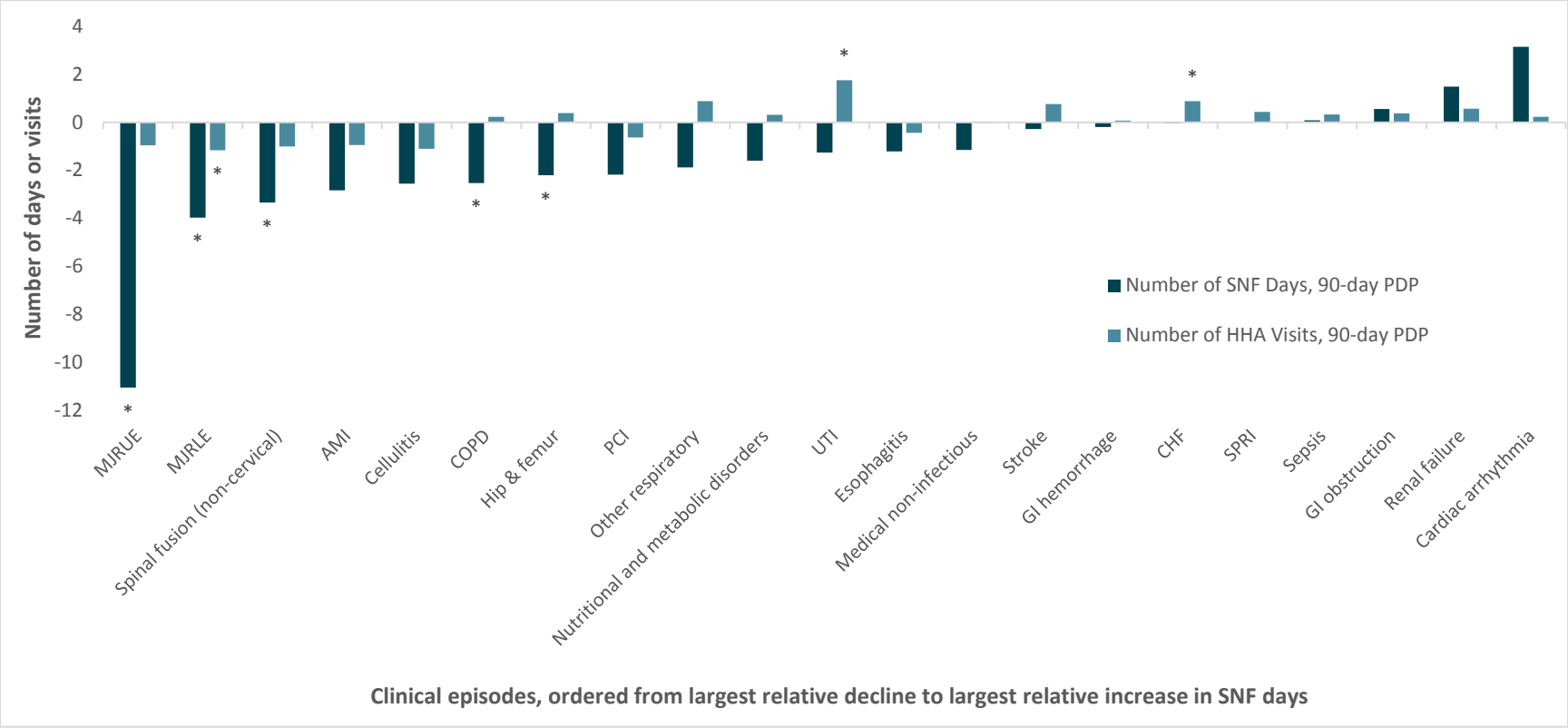


Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. PAC = post-acute care. PGP= physician group practice. MJRUE=major joint replacement of the upper extremity. GI=gastrointestinal. MJRLE=major joint replacement of the lower extremity. UTI=urinary tract infection. CHF=congestive heart failure. COPD=chronic obstructive pulmonary disease. SPRI=simple pneumonia and respiratory infections. AMI=acute myocardial infarction. PCI=percutaneous coronary intervention.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 37: Impact of BPCI on SNF Days and HHA Visits, by Clinical Episode, Model 2 PGP, Baseline to Intervention, Q4 2013 – Q4 2016



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model ordered starting with the largest relative decline in the number of SNF days, 90-day PDP. These utilization measures are conditional upon use of the service. Beneficiaries must have spent a minimum of one day in a skilled nursing facility (SNF) and had at least one home health agency (HHA) visit during the 90 day post-discharge period (PDP) to be included in the DiD estimate for number of SNF days and HHA visits, respectively. PGP= physician group practice. MJRUE=major joint replacement of the upper extremity. MJRLE=major joint replacement of the lower extremity. AMI=acute myocardial infarction. COPD=chronic obstructive pulmonary disease. PCI=percutaneous coronary intervention. UTI=urinary tract infection. GI=gastrointestinal. CHF=congestive heart failure. SPRI=simple pneumonia and respiratory infections.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

How has quality of care changed under BPCI?

Claim-based results suggest that the quality of care generally did not change under BPCI Model 2 for PGP-initiated episodes (Exhibit 38). The all-cause mortality rate during the 90-day PDP for episodes associated with BPCI-participating PGPs declined relative to comparison episodes from similar hospitals in 10 of the 21 clinical episodes; the decline was statistically significant in two clinical episodes ($p<0.05$). There were also statistically significant relative increases in the mortality rate for two BPCI PGP clinical episodes: hip and femur procedures except major joint and nutritional and metabolic disorders ($p<0.10$). Data from the baseline period shows that hip and femur episodes associated with BPCI PGPs and matched episodes from similar hospitals were not on parallel trends for mortality rates, which is required for an unbiased estimate. For nutritional and metabolic disorders, the relative increase in mortality for BPCI PGP episodes was associated with increased hospice use. There was an increase in the proportion of BPCI PGP episodes that used hospice from baseline to the intervention period relative to the change for the comparison group. In addition, controlling for hospice use in the risk-adjustment regression model resulted in a DiD estimate that was smaller, though still statistically significant ($p<0.10$).

Emergency department use during the 90-day PDP for episodes associated with BPCI-participating PGPs had statistically significant relative declines for two clinical episodes ($p<0.05$). The sensitivity test result suggests that the decrease for percutaneous coronary intervention episodes may be due to the random selection of episodes in the matched sample, and it may not have been statistically significant if a different sample of comparison episodes had been selected. (Results of the sensitivity test are presented in **Appendix H**.)

There were statistically significant relative declines in 90-day readmission rates for two clinical episodes ($p<0.10$). However, the results from the sensitivity test for MJRLE and nutritional and metabolic disorders suggest that the results may be due to the random comparison episode selection, and they may not have been statistically significant if a different sample of comparison episodes had been selected (see **Appendix H** for results). In addition, data from the baseline period for nutritional and metabolic disorders shows that episodes from BPCI PGPs and comparison episodes from hospitals similar to those associated with BPCI PGPs were not on parallel trends, which is required for an unbiased estimate. There was a statistically significant relative increase in readmission rates for three clinical episodes ($p<0.10$). The relative increase in readmissions for COPD, bronchitis, asthma episodes may be due to the particular baseline quarters (Q4 2011 through Q3 2012) in the DiD analysis, suggesting the change was not necessarily a result of BPCI.[38]

[38] To determine whether the baseline period was by chance a year with a low readmission rate, we calculated the DiD estimates using a longer baseline period (Q4 2010 through Q3 2013) using the same comparison group of providers. The three year baseline time period results in a risk-adjusted DiD estimate that is not statistically significant.

Exhibit 38: Impact of BPCI on Claim-based Quality Outcomes, by Clinical Episode, Model 2 PGP, Baseline to Intervention, Q4 2013-Q4 2016

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	All-cause Mortality Rate, 90-day PDP (pp)	ED Use, 90-day PDP (pp)	Unplanned Readmission Rate, 90-day PDP (pp)
Acute myocardial infarction	3,772	1.5	-1.6	0.6
Cardiac arrhythmia	2,978	1.6*	1.7	**3.5**
Cellulitis	2,859	**-2.3**	-0.7	2.5
Congestive heart failure	9,584	-1.2*	0.2	0.4
COPD, bronchitis, asthma	9,200	0.2	0.3*	2.0
Esophagitis, gastroenteritis & other digestive disorders	3,723	0.2	1.4	0.5
Gastrointestinal hemorrhage	2,461	0.4	-0.3	-1.0
Gastrointestinal obstruction	1,563	-0.1*	-2.1	-0.7
Hip & femur procedures except major joint	6,328	1.5*	-0.5	0.7
Major joint replacement of the lower extremity	64,392	-0.1*	**-1.0**	-0.5^
Major joint replacement of the upper extremity	2,476	0.1	-0.4	0.1
Medical non-infectious orthopedic	3,437	-0.4	-1.1*	-1.5
Nutritional and metabolic disorders	2,674	**3.2**	1.6	-3.3*
Other respiratory	3,742	-1.8	0.6*	-0.1
Percutaneous coronary intervention	2,722	-0.1	**-3.8*^**	-1.4
Renal failure	5,899	-1.2	0.0	**2.7**
Sepsis	20,802	0.3	-0.2	-0.7
Simple pneumonia & respiratory infections	10,117	1.0	-0.1	-0.4
Spinal fusion (non-cervical)	2,391	-0.5	-2.7	-1.4
Stroke	3,485	0.3	-1.6	0.2
Urinary tract infection	6,422	**-2.4**	-0.2	0.1

Note: The estimates in this table are the results of a difference-in-differences model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented due to insufficient sample size. PGP=physician group practice. PDP=post-discharge period. pp=percentage points. ED=Emergency Department. COPD=chronic obstructive pulmonary disease.

* This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate. Equal trends test was conducted for total allowed payment amount and total amount paid by Medicare IP through 90-day PDP, emergency department visits, readmission, and mortality outcomes.

^ The sensitivity test suggests that the statistically significant result may be due to the random selection of comparison episodes in the matched sample. See **Appendix D** for additional information on the sensitivity test methodology. Results of the sensitivity test are presented with the BPCI impact estimates in **Appendix H**.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

2. Patient Functional Status and Health Care Experience

This section presents patient care experiences and functional outcomes from the beneficiary survey from February 2017 through September 2017.

a. Key Findings

- There was no consistent difference between BPCI and comparison respondents with regard to changes in self-reported functional status from before to after the anchor hospitalization.
- BPCI respondents reported statistically significantly worse care experience for three of nine measures relative to the comparison group. However, there was no statistically significant difference between BPCI and comparison respondents in the proportion satisfied with their overall recovery since leaving the hospital.

b. Methods

The beneficiary survey used a cross-sectional regression approach to estimate risk-adjusted differences in self-reported patient outcomes between beneficiaries who received services from BPCI EIs and those who received services from a comparison group of non-BPCI providers, during the BPCI intervention period. The survey collected information from beneficiaries about seven measures of change in functional status (improvement or decline) from before to after hospitalization, nine measures of care experience, and a measure of overall satisfaction with physical recovery.

The survey sample for BPCI Model 2 with physician group practice (PGP) EIs contains 18 out of the 48 eligible clinical episodes. CMS defines a Model 2 PGP episode as having both an inpatient claim with the attending physician billing under the PGP's TIN, and at least one post-discharge visit with an NPI also billing under the same PGP's TIN. Our survey sampling did not permit delay (run-out time) to await post-discharge visit claims. We therefore sampled based only on the inpatient attending physician's PGP TIN, without regard to any subsequent visit claim with the same PGP TIN.[39] To the extent that some BPCI PGP survey respondents did not receive any post-discharge care from the PGP because of this sampling approach, estimates may not reflect complete BPCI episodes as defined by CMS.

BPCI and comparison beneficiaries selected for the Model 2 PGP survey sample were matched by age, presence of major complication or comorbidity (MCC), and hospital size. Clinical episodes with sufficient volume were also matched on hospital academic affiliation.[40] Data were collected across three waves covering February 2017 through September 2017.[41] Responses from all clinical episodes and all waves were pooled together for analysis.

Across all waves and clinical episodes, the response rate averaged 45.3% and varied considerably for the different clinical episodes, ranging from 71.6% to 36.0%. We applied non-response and

[39] We included episodes in this analysis if a patient had contact with a BPCI NPI in the hospital (and was found on a claim for the episode), even if the episode was not ultimately attributed to a BPCI-participating PGP.
[40] PGP Comparison episodes were defined using the same criteria as hospital comparison episodes. Therefore, PGP episodes were matched based on the characteristics of the discharging hospital.
[41] For the exact months covered by each wave of the survey, see **Appendix D**.

sampling weights to all observations. Estimated differences between the BPCI and comparison respondents were risk-adjusted for beneficiary- and hospital-level characteristics. Since survey data were only collected during the intervention period, we can identify differences between BPCI and comparison respondents, but cannot determine whether such differences are attributable to BPCI or other, pre-existing factors.

See **Appendix D** for additional details about the survey measures, sample selection, and other methods.

c. Results

The survey results are based on 16,898 BPCI and 14,652 comparison responses.[42] The 18 clinical episodes included in the analysis represent 89% of all BPCI Model 2 PGP episode volume since the start of the initiative. This analysis, therefore, approximates results across all Model 2 PGP clinical episodes from February 2017 through September 2017.

Overall, BPCI had no meaningful association with self-reported changes in functional status from before to after the anchor hospitalization (Exhibit 39). Differences in the rate of improvement and decline were not statistically significant for five out of seven functional status measures; point estimates were small and did not follow any consistent pattern. BPCI respondents were 1.15 percentage points more likely to report improvement in using stairs, relative to comparison respondents (BPCI 45.00 vs. 43.84 comparison; $p<0.10$), suggesting improved functional status. However, BPCI respondents were also 1.03 percentage points more likely to report increased use of a mobility device relative to comparison respondents (BPCI 36.60 vs. 35.56 comparison; $p<0.10$), which may suggest diminished functional status. Since these two results are inconsistent, and all other results are not statistically significant, there is no evidence that BPCI and comparison respondents differed in terms of functional change from before to after their episodes.

Exhibit 39: Differences in Survey-based Quality Outcomes between BPCI and Comparison Respondents, Model 2 PGP, February 2017 – September 2017

Domain	Survey Measure	BPCI Rate (%)	Comparison Rate (%)	Difference (pp)
Changes in Functional status	Improvement in bathing, dressing, using toilet, or eating	72.77	72.00	0.77
	Decline in bathing, dressing, using toilet, or eating	14.43	14.79	-0.36
	Improvement in planning regular tasks	61.59	61.51	0.08
	Decline in planning regular tasks	21.92	22.05	-0.13
	Improvement in use of a mobility device (less likely to use)	49.94	50.41	-0.48
	Decline in use of a mobility device (more likely to use)	36.60	35.56	1.03
	Improvement in walking without rest	45.27	45.01	0.26
	Decline in walking without rest	27.14	27.40	-0.26
	Improvement in using stairs	45.00	43.84	1.15
	Decline in using stairs	30.57	31.32	-0.75
	Physical/emotional problems limit social activities less frequently	59.66	60.52	-0.86
	Physical/emotional problems limit social activities more frequently	22.33	21.61	0.72
	Pain limits regular activities less frequently	57.26	57.31	-0.05
	Pain limits regular activities more frequently	19.76	19.38	0.38

[42] We oversampled BPCI episodes expecting that many of the episodes ultimately would be dropped from the analysis. However, due to the analytic method we selected, more BPCI episodes remained in the final sample.

Domain	Survey Measure	BPCI Rate (%)	Comparison Rate (%)	Difference (pp)
Care Experience	Never received conflicting medical advice	72.82	73.73	-0.92
	Services always appropriate for level of care patient needed	61.89	63.71	-1.82
	Medical staff always spoke in patient's preferred language	93.91	93.49	0.42
	Agree that patient was discharged at the right time	88.50	89.66	-1.16
	Agree that medical staff took patient's preferences into account in deciding post-discharge health care services	92.02	92.92	-0.90
	Agree that patient had good understanding of how to take care of self before going home	94.58	94.85	-0.26
	Agree that medical staff clearly explained how to take medications before going home	94.10	94.23	-0.14
	Agree that medical staff clearly explained what follow-up appointments would be needed before patient went home	94.79	94.80	-0.01
	Agree that patient had been able to manage health needs since returning home	95.88	96.10	-0.22
Overall Satisfaction	Extremely or quite a bit satisfied with overall satisfaction with recovery since leaving the hospital	70.37	71.12	-0.75

Notes: The estimates in this table are the result of a cross-sectional logistic regression risk adjustment model for binary indicators. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Estimates were based on 16,898 BPCI survey respondents and 14,652 comparison survey respondents. All responses were weighted for non-response and sampling design. The analysis pooled all 18 Model 2 PGP clinical episodes that were analyzed at the episode level, across survey waves 9 through 11. pp = percentage points.

***Source*:** Lewin analysis of BPCI Beneficiary Survey responses for episodes that began in February 2017 through September 2017.

The proportion of respondents with favorable care experiences was slightly smaller for BPCI survey respondents than comparison respondents on three of the nine measures of care experience. Relative to comparison respondents, BPCI respondents were less likely to report that their level of care was always appropriate (-1.82 percentage points: 61.89 BPCI vs. 63.71 comparison; $p<0.05$), less likely to say they were discharged at the right time (-1.16 percentage points: 88.50 BPCI vs. 89.66 comparison; $p<0.05$), and less likely to agree that medical staff took their preferences into account in deciding what health services they should have after leaving the hospital (-0.90 percentage points: 92.02 BPCI vs. 92.92 comparison; $p<0.05$). The difference in the proportion of respondents indicating that they were "quite a bit" or "extremely" satisfied between the BPCI and comparison groups was also negative, but it was small and not statistically significant.

Although BPCI had a negative and statistically significant association with three measures of care experience, the differences were small (ranging from -0.90 to -1.82 percentage points), and there were no systematic differences in functional status outcomes.

Appendix I presents the results by clinical episode. Results from 11 of the 18 clinical episodes indicated no difference in functional status between BPCI and comparison respondents. Among the seven clinical episodes where there were consistent differences in functional status, there were four in which BPCI respondents reported better outcomes than comparison respondents, and three in which BPCI respondents reported worse outcomes.

The slightly less favorable care experiences reported by BPCI respondents at the aggregate level cover a substantial amount of heterogeneity of results across clinical episodes. There were no consistent patterns of differences between BPCI and comparison respondents in 10 out of 18

clinical episodes. BPCI respondents indicated more favorable care experience outcomes in three out of 18 clinical episodes and less favorable care experiences in five out of 18 clinical episodes.

3. Understanding the Impact of Terminating Participation on Payments

The BPCI initiative allows participants to withdraw from a clinical episode at any time.[43] By December 2016, 54% of the Model 2 PGP EIs withdrew from at least one clinical episode.[44] Descriptive analysis suggests that PGPs were more likely to withdraw from clinical episodes in which they had negative NPRA, so withdrawal from BPCI is non-random. This non-random withdrawal could affect our BPCI impact estimates by disrupting the balance between the BPCI and comparison groups, as well as giving more weight to participants with a longer exposure to BPCI. It is not clear how these potential biases would affect the BPCI impact estimates. To evaluate the impact of the withdrawals on the impact estimates, we re-calculated the estimates with all episodes from participants and comparison group providers included in the analysis through December 2016, regardless of whether the BPCI EI had withdrawn from the clinical episode prior to December 2016. We refer to these sensitivity analyses as ITT DiD analyses.

Across all clinical episodes, 7% of the matched PGP EI episodes were initiated after the EI had withdrawn from that clinical episode. The share of matched post-withdrawal episodes as a proportion of the intervention sample varied widely across clinical episodes, from 3% for MJRLE to 13% for major joint replacement of the upper extremity and spinal fusion (non-cervical). Results by clinical episode are located in **Appendix K**.

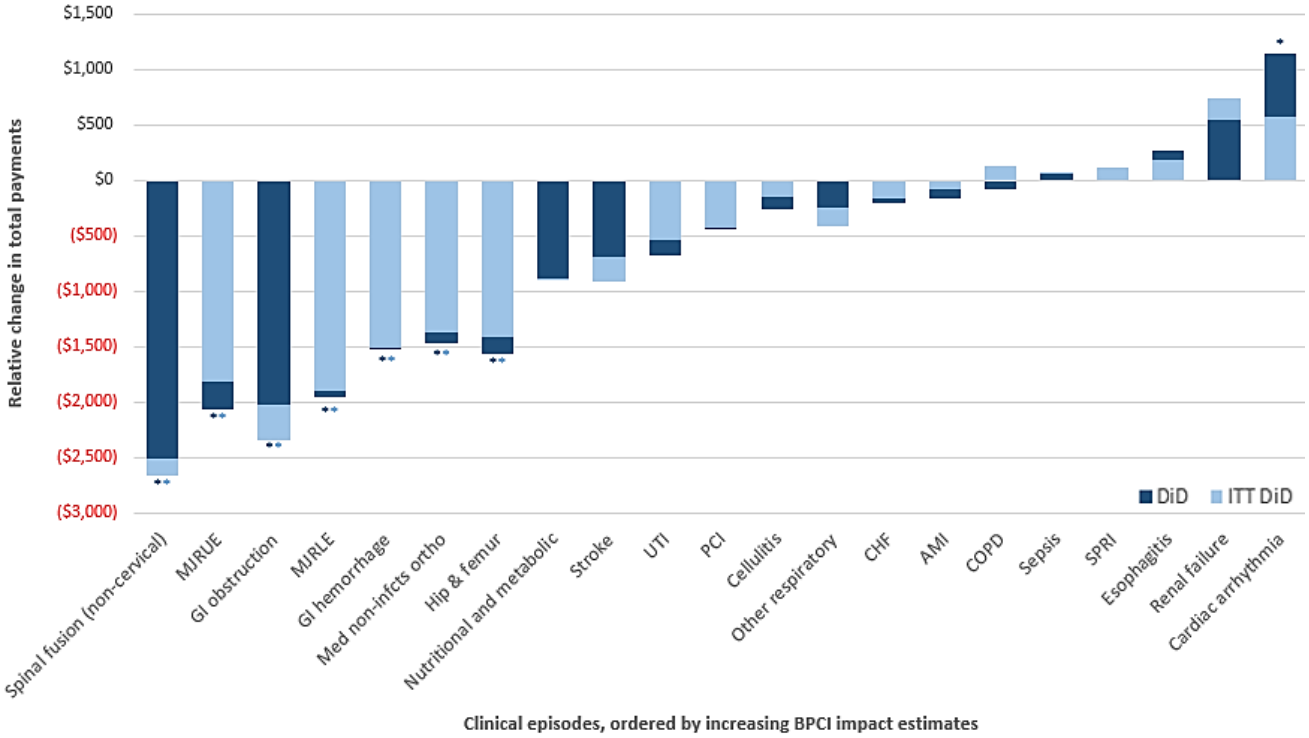
All seven of the Model 2 PGP clinical episodes that exhibited a statistically significant relative decline in total allowed payments also had a statistically significant ITT DiD estimate in the same direction.[45] The one clinical episode that had a statistically significant relative increase in total allowed payments also exhibited a relative increase in the ITT DiD estimate, but it was not statistically significant (Exhibit 40). This indicates that despite selective withdrawal, BPCI participants achieved a decline in total allowed payment amounts relative to the comparison group for these clinical episodes. For additional results and methods, see **Appendix K**.

[43] We use the term withdrawal to reflect EIs that terminate their participation in BPCI or stop participating in the given clinical episode. Note that EIs that participated in a clinical episode for only one quarter are excluded from both the original and the ITT DiD analysis.

[44] This is limited to participants that met the minimum number of patient episodes to be included in our DiD and ITT DiD analysis.

[45] Furthermore, five of the seven clinical episodes that exhibited significant relative declines in the original impact estimate exhibited smaller relative declines in the ITT DiD estimate. This suggests that in certain cases the BPCI effect may have diminished after withdrawal.

Exhibit 40: BPCI Impact Estimates and ITT DiD Estimates of Total Medicare Allowed Payments during the Inpatient Stay and 90-days Post discharge, by Clinical Episode, Model 2 PGP, Baseline to Intervention, Q4 2013 – Q4 2016



Note: The estimates in this exhibit are the results of the BPCI impact difference-in-differences (DiD) model (dark blue) and intent-to-treat (ITT) DiD model (light blue). The results are ranked by increasing magnitude of the BPCI impact estimates. BPCI impact estimates that are significant at the 10% significance level are indicated by a dark blue asterisk; ITT DiD estimates that are significant at the 10% significance level are indicated by a light blue asterisk. The Medicare payment outcome is standardized to remove the effect of geographic and other adjustments. PGP=physician group practice. MJRUE=major joint replacement of the upper extremity. GI=gastrointestinal. MJRLE=major joint replacement of the lower extremity. UTI=urinary tract infection. PCI=percutaneous coronary intervention. CHF=congestive heart failure. AMI=acute myocardial infarction. COPD=chronic obstructive pulmonary disease. SPRI=simple pneumonia and respiratory infections.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

4. Market Dynamics

This section presents the analysis of market share and provider referral patterns for Model 2 PGPs in the MJRLE clinical episode that entered Phase 2 of BPCI in Q3 2015.

a. Key Findings

- There was a small increase in the market share of MJRLE episodes from the baseline period to the BPCI period among M2 PGP EIs that entered Phase 2 of BPCI in Q3 2015.
- This cohort of Model 2 PGP EIs did not tend to concentrate their discharges to a smaller group of PAC providers (SNFs or HHAs) while participating in BPCI, relative to the baseline period.
- There was a moderate increase in the share of Model 2 PGP MJRLE patients who were discharged to highly rated SNFs in the intervention period compared to the baseline, but there was no change in the share of patients who were discharged to highly rated HHAs.

b. Methods

We used the same analytical approach that was used to analyze Model 2 hospital EIs described in Section II.B.5 above.[46] In addition to analysis stratified by six-month intervals, we also calculated market measure rates by combining the latest two six-month intervals in the BPCI period (Q1 2016 through Q4 2016; referred to as "year 2016" thereafter) and compared them to the baseline period. Details about the analytical methods are in **Appendix D**.

c. Results

The median MJRLE market share among Model 2 PGP EIs that entered Phase 2 in Q3 2015 was 4.9% in the baseline period. There was a median increase of 0.3 pp from the baseline period to the year 2016 (Exhibit 41), which is consistent with increases in market share across the three six-month intervals of the BPCI period (see **Appendix L** for complete results).[47]

In the baseline period, PGP EIs discharged 75% of their MJRLE patients to a median of six SNF providers and three HHA providers. There was no change in the number of SNF or HHA providers used from the baseline period to the year 2016 (Exhibit 41). There was also no consistent change in the number of SNF or HHA providers across the three six-month intervals in the BPCI period (see **Appendix L** for results).

PGP EIs discharged a median of 32.6% of their MJRLE patients to highly rated SNF providers and 21.9% to highly rated HHA providers. There was virtually no change in the share of Model 2 PGP MJRLE patients who were discharged to highly rated HHAs in year 2016, but there was a moderate increase of 4.0pp in the share of Model 2 PGP MJRLE patients who were discharged to highly rated SNFs (Exhibit 41). Model 2 PGP EIs also tended to discharge more MJRLE patients to highly rated SNFs in every six-month interval of the BPCI period (see **Appendix L** for results).

Exhibit 41: Median Baseline Rate and Median Changes from Baseline for Market Dynamics Measures, Model 2 PGP Major Joint Replacement of the Lower Extremity EIs, Q4 2011- Q4 2016

Measures	Median Baseline Rate	Median Changes from Baseline
EI Market Share (%)	4.9	0.3
Number of PAC Providers Used: SNFs	6.0	0.0
Number of PAC Providers Used: HHAs	3.0	0.0
Percent of PAC Admissions to High Star-Rating SNFs (%)	32.6	4.0
Percent of PAC Admissions to High Star-Rating HHAs (%)	21.9	0.03

Note: This table shows the measure rate in the baseline period, and the changes from the baseline period to the last BPCI year in the analysis (Q1 2016 through Q4 2016 combined). PGP=physician group practice. EI=episode initiator. SNF=skilled nursing facility. HHA=home health agency. PAC=post-acute care.

Source: Lewin analysis of Medicare claims for discharges that began Q4 2011 through Q4 2016.

[46] We assigned a hospital EI to the market (i.e., a Medicare CBSA) where the EI is physically located. The market assignment for PGP EIs is more complex because PGPs may have a presence in multiple markets. For PGP EIs, we calculated market measures at the PGP EI and market level, including all markets where a PGP EI had an inpatient discharge.

[47] Consistent change in this section means non-zero changes in the same direction from the baseline to each six-month interval in the intervention period, regardless of the magnitude of the changes.

D. Factors Contributing to the Variation in NPRA among BPCI Providers

This section examines the factors that may have contributed to the variation in the financial performance of BPCI-participating hospitals and PGPs. The goals of this analysis were to: distinguish the provider, model, and market characteristics of top and bottom BPCI performers; identify strategies top performers employ to decrease Medicare payments; and assess whether there are unintended consequences correlated with performance in the BPCI initiative. Unlike the BPCI impact analysis, which uses a DiD approach and quantifies the overall impact of BPCI by comparing changes in outcomes of BPCI providers to changes of a comparison group, this analysis assesses variation among BPCI hospitals and PGPs. Furthermore, whereas the BPCI impact analysis uses the total allowed payment amount to assess financial performance, this analysis focuses on providers' standardized net payment reconciliation amounts (NPRA) (henceforth "NPRA").[48] NPRA measures the difference between the target price for services provided during the episode of care and the total dollar amount of Medicare fee-for-service expenditures for that episode.[49] We examined financial performance as measured by NPRA for 14 hospital clinical episodes and five PGP clinical episodes in Model 2.[50]

1. Key Findings

- On average, higher performing EIs (i.e., hospitals and PGPs with the largest average NPRA per episode relative to their target price) across clinical episodes reduced their use of institutional PAC, increased the share of patients discharged home without any PAC, and reduced the rate of unplanned readmissions, relative to lower performing EIs.
- Circumstances prior to the start of BPCI may have made it easier for hospitals and PGPs with higher standardized NPRA to reduce their use of institutional PAC during the intervention as they discharged a larger share of their patients to PAC during the baseline period.
- While higher performers made the largest shifts towards less expensive PAC services, they also experienced the most favorable shifts in patient mix, as evidenced by the relative reductions in the average HCC score and age of their patients.

2. Methods

Multiple data sources were used for this analysis. First, we used episode files from the reconciliation contractor for the NPRA from Q4 2013 through Q4 2016. Independent variables in our analysis came from the following data sources: Medicare claims data (Q4 2011 through Q3 2012 for the baseline and Q4 2013 through Q4 2016 for the intervention period), the CMS BPCI

[48] Standardized NPRA is NPRA expressed as a percentage of the provider's target price for a given Model and clinical episode.

[49] When a provider's episode payments are below the target price (i.e., positive NPRA), the provider receives this amount from CMS. When a provider's episode payments exceed the target price (i.e., negative NPRA) the provider may need to return the amount to CMS. The target price is set by discounting historical episode payments by 2 or 3 percent, depending on the Model and episode options selected.

[50] BPCI providers were required to meet certain criteria to be included in this analysis such as having at least 50 episodes during both the baseline and intervention periods. For more detail on the inclusion and exclusion criteria, please see the methods section below.

database (Q4 2013 through Q4 2016), implementation protocol data (Q4 2013 through Q4 2016), and data collected from Awardees (Q4 2013 through Q4 2015).

To identify what characterizes top and bottom performers under the BPCI initiative, we assigned BPCI EIs into mutually exclusive performance groups based on standardized NPRA from the first quarter they joined BPCI through Q4 2016. Standardized NPRA measures a provider's average NPRA as a percent of the provider's target price for a given Model and clinical episode.[51] We selected this metric for two reasons. First, it assumes that participants repaid any negative NPRA through the entire intervention period,[52] which allows us to measure the potential monetary gains received by providers. Second, it does not favor EIs that had the highest savings opportunities because they had the highest payments during the baseline period. We calculated standardized NPRA as follows,

$$Standardized\ NPRA = \frac{(Average\ EI\ NPRA)}{(Weighted\ Average\ EI\ Target\ Price)}$$

where *Average EI NPRA* is the average NPRA per episode and it is defined as:

$$\frac{Q_1 NPRA + Q_2 NPRA + \cdots Q_N NPRA}{\sum_{i=1}^{N} EI\ Cases\ in\ Q_i}$$

and *Weighted Average EI Target Price* is the average target price, weighted by the number of episodes for each, and it is calculated as:

$$\frac{[(Q_1 EI\ Target\ Price\ \times Q_1 EI\ Cases) + (Q_2 EI\ Target\ Price\ \times Q_2 EI\ Cases) + \cdots (Q_i EI\ Target\ Price\ \times Q_N EI\ Cases)]}{\sum_{i=1}^{N} EI\ Cases\ in\ Q_i}$$

where $Q_i NPRA$ represents a provider's NPRA in quarter i, $Q_i EI\ Target\ Price$ represents a provider's target price for the given clinical episode in quarter i, and $EI\ Cases\ in\ Q_i$ represents the number of cases in quarter i. For a given clinical episode, top performers were defined as participants with standardized NPRA above the 75th percentile of the standardized NPRA distribution. Average performers were defined as participants with standardized NPRA between the 26th and 75th percentiles of the standardized NPRA distribution. Bottom performers were those with standardized NPRA at or below the 25th percentile of the distribution.

To ensure meaningful results, we only included combinations of Model/EI type/clinical episodes with sufficient volume for the analyses. To meet our inclusion criteria, each combination must have had at least 30 BPCI EIs with at least 50 episodes each during the baseline and intervention periods with a non-missing value for the total Medicare standardized allowed payment for the inpatient stay plus 90 days post-discharge outcome. We analyzed 14 Model 2 hospital and five Model 2 PGP clinical episodes that met our selection criteria. No Model 3 clinical episodes met our selection criteria.

[51] To account for the changes in target price and variation in the number of episodes each quarter, we estimated a weighted average target price, weighted by the number of episodes.

[52] CMS eliminated downside risk from Q4 2013 through Q4 2014. They also eliminated negative NPRA for any episode of care that was initiated as a result of the episode attribution issues caused by the incorrect PGP Reassignment Lists for the period of January 1, 2015 through September 30, 2016.

We conducted descriptive analyses, including frequencies and other univariate statistics, to examine distributions of our performance measure and independent variables during the baseline and intervention periods. The summary statistics reflect how the performance of each group (i.e., top performers, average performers, bottom performers) correlates with a variety of characteristics of interest. Specifically, we correlated performance with model, provider, and market characteristics; payment, utilization, and quality outcomes during the baseline period; patient mix during the baseline period; and shifts in payment, utilization, quality, and patient mix between the baseline and the intervention periods.

To assess whether there were meaningful correlations, we used Spearman's Rank Order Correlation. This nonparametric test measures the strength and direction of the association between two ranked variables, in this case, our performance measure (i.e., standardized NPRA) and the characteristic of interest. The correlation is less sensitive to outliers as it calculates the correlation on ranked values rather than on the actual values. Spearman's coefficient (rho) ranges from -1 to +1. Values close to +/- 1 indicate a stronger positive/negative association between two variables; a value of 1, for example, means a perfect positive association of ranks. Values close to zero signify a weak association between variables.

3. Results

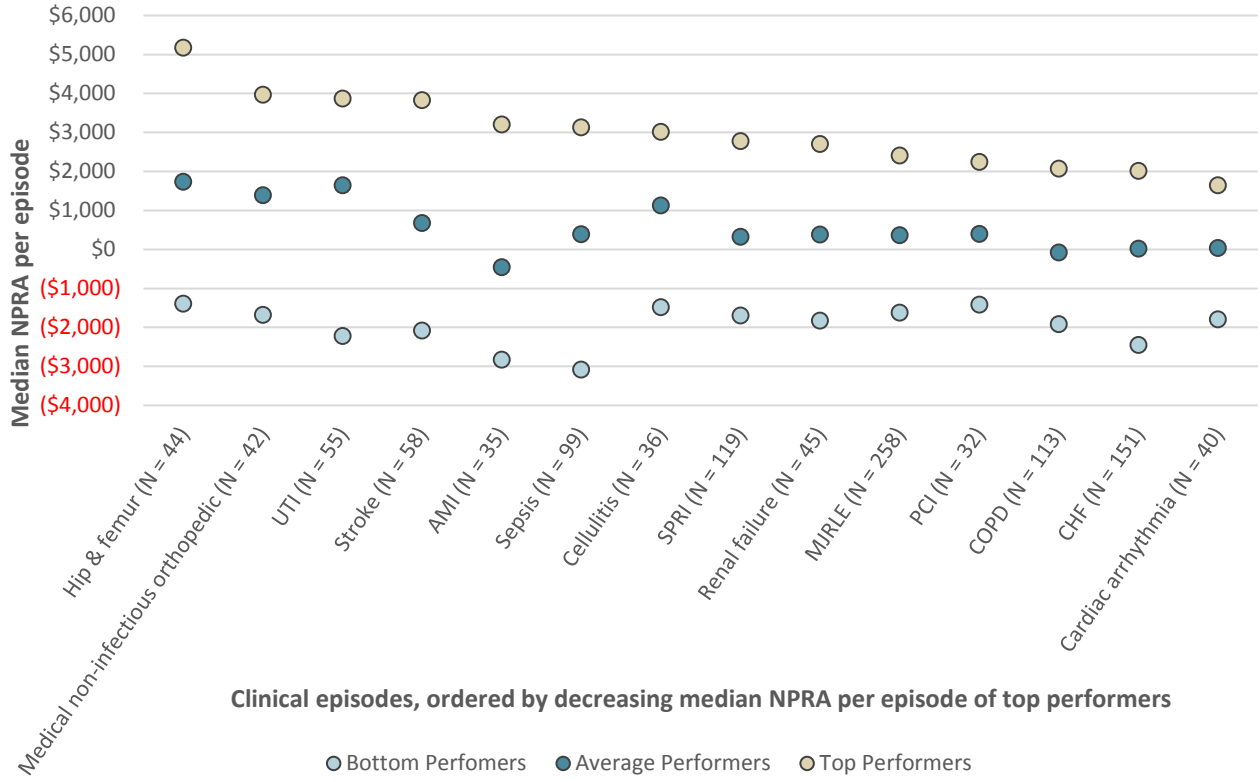
a. BPCI-participating Hospitals

Differences in Performance across Participating Hospitals

Across the 14 clinical episodes we analyzed, we found variation in financial performance, as measured by NPRA among BPCI hospitals, as shown in Exhibit 42. Top performers in each clinical episode had a median NPRA that was between $3,439 and $6,560 more than that of bottom performers in the same clinical episode. The clinical episode with the largest difference in median NPRA between top and bottom performers was hip and femur procedures except major joint. Top performing hospitals participating in this episode had a median NPRA of $5,174, while bottom performers had a median NPRA of -$1,386. Conversely, the clinical episode with the smallest difference in median NPRA between top and bottom performers was cardiac arrhythmia, with a median NPRA of $1,649 among top performers and -$1,790 among bottom performers.

We also found variation in financial performance across clinical episodes for each performance group. Across the 14 clinical episodes, the median NPRA for top performers ranged from $1,649 to $5,174. Among average performers, the median NPRA ranged from -$457 to $1,738. Finally, among bottom performers, the median NPRA ranged from -$3,082 to -$1,386. Compared to other clinical episodes, hospitals participating in the hip and femur except major joint clinical episode had the largest median gains among top and average performers ($5,174 and $1,738, respectively). Hospitals in this clinical episode also had the smallest median loss among bottom performers (-$1,386), whereas bottom performers in the sepsis clinical episode had the largest median loss (-$3,082).

Exhibit 42: Median NPRA per Episode, by Performance Group, Model 2 Hospitals, Q4 2013 – Q4 2016



Note: NPRA = net payment reconciliation amount. UTI = urinary tract infection. AMI = acute myocardial infarction. SPRI = simple pneumonia and respiratory infections. MJRLE = major joint replacement of the lower extremity. PCI = percutaneous coronary intervention. COPD = chronic obstructive pulmonary disease. CHF = congestive heart failure.

Source: Lewin analysis of net payment reconciliation amount data for episodes that began Q4 2013 through Q4 2016.

Correlation between NPRA and Stopping Participation in a Clinical Episode

Across clinical episodes, the financial performance of hospitals is consistent with the proportion of hospital EIs that stopped participating in the clinical episode. Through Q4 2016, top performing hospitals had the lowest dropped episode rate for 12 out of 14 episodes, while bottom performers had the highest dropped episode rate for 12 out of the 14 episodes. Across clinical episodes, the dropout rate ranged from 0% to 40% for top performers, 6% to 55% for average performers, and 23% to 89% for bottom performers, as shown in Exhibit 43. Overall, for all performance groups, the percutaneous coronary intervention clinical episode had the lowest share of EIs that stopped participating (16%), while cellulitis had the highest (47%).

Exhibit 43: Episode Initiators that Stopped Participating in the Clinical Episode, by Performance Group, Model 2 Hospitals, Q4 2013 – Q4 2016

Clinical Episode	Performance Group	EIs (N)	Number that Stopped Participating	Percent that Stopped Participating	Median NPRA per Episode
Acute myocardial infarction	Top Performers	8	1	13%	$3,207
	Average Performers	18	9	50%	-$457
	Bottom Performers	9	4	44%	-$2,822
	Total	35	14	40%	-$488
Cardiac arrhythmia	Top Performers	10	4	40%	$1,649
	Average Performers	20	11	55%	$43
	Bottom Performers	10	3	30%	-$1,790
	Total	40	18	45%	$43
Cellulitis	Top Performers	9	0	0%	$3,018
	Average Performers	18	9	50%	$1,133
	Bottom Performers	9	8	89%	-$1,479
	Total	36	17	47%	$1,133
Congestive heart failure	Top Performers	37	6	16%	$2,018
	Average Performers	76	19	25%	$25
	Bottom Performers	38	20	53%	-$2,444
	Total	151	45	30%	$10
COPD, bronchitis, asthma	Top Performers	28	1	4%	$2,076
	Average Performers	56	18	32%	-$78
	Bottom Performers	29	15	52%	-$1,912
	Total	113	34	30%	-$106
Hip and femur procedures except major joint	Top Performers	11	1	9%	$5,174
	Average Performers	22	6	27%	$1,738
	Bottom Performers	11	7	64%	-$1,386
	Total	44	14	32%	$1,738
Medical non-infectious orthopedic	Top Performers	10	0	0%	$3,974
	Average Performers	21	6	29%	$1,395
	Bottom Performers	11	9	82%	-$1,679
	Total	42	15	36%	$1,362
Major joint replacement of the lower extremity	Top Performers	64	9	14%	$2,410
	Average Performers	129	22	17%	$369
	Bottom Performers	65	15	23%	-$1,618
	Total	258	46	18%	$363
Percutaneous coronary intervention	Top Performers	8	1	13%	$2,244
	Average Performers	16	1	6%	$397
	Bottom Performers	8	3	38%	-$1,410
	Total	32	5	16%	$397

Clinical Episode	Performance Group	EIs (N)	Number that Stopped Participating	Percent that Stopped Participating	Median NPRA per Episode
Renal failure	Top Performers	11	1	9%	$2,710
	Average Performers	22	9	41%	$388
	Bottom Performers	12	6	50%	-$1,825
	Total	45	16	36%	$352
Sepsis	Top Performers	24	1	4%	$3,133
	Average Performers	50	23	46%	$393
	Bottom Performers	25	20	80%	-$3,082
	Total	99	44	44%	$384
Simple pneumonia and respiratory infections	Top Performers	29	1	3%	$2,778
	Average Performers	60	18	30%	$329
	Bottom Performers	30	12	40%	-$1,689
	Total	119	31	26%	$324
Stroke	Top Performers	14	0	0%	$3,829
	Average Performers	29	7	24%	$679
	Bottom Performers	15	8	53%	-$2,075
	Total	58	15	26%	$636
Urinary tract infection	Top Performers	13	2	15%	$3,871
	Average Performers	28	7	25%	$1,650
	Bottom Performers	14	7	50%	-$2,214
	Total	55	16	29%	$1,646

Note: Median NPRA per episode is representative of all providers within a clinical episode performance group and is not limited to only providers who stopped participation. This table includes hospitals that stopped participating in the clinical episode for any reason, including withdrawing from BPCI. NPRA = net payment reconciliation amount. COPD = chronic obstructive pulmonary disease.

Source: Lewin analysis of CMS' BPCI database and net payment reconciliation amount data for episodes that began Q4 2013 through Q4 2016.

Differences in Model, Hospital, and Market Characteristics across Performance Groups

We analyzed descriptive statistics for 12 model, hospital, and market characteristics by performance group to determine whether high performance was associated with these characteristics. We ran a Spearman rank-order correlation on six out of the 12 characteristics to analyze whether higher performance was statistically correlated with the characteristic (Exhibit 44).[53] Exhibit 45 presents the remaining model, hospital, and market characteristics by performance group.

BPCI affords participants flexibility in how to participate in the model and we observed associations between higher performing hospitals and the model features they selected. Higher performance was associated with selecting a larger number of clinical episodes to participate in for 12 out of 14 clinical episodes (Exhibit 44). This association was statistically significant for two clinical episodes ($p<0.05$). In half of the clinical episodes, top performers were less likely to have

[53] We were unable to conduct a Spearman rank-order correlation for the remaining six binary characteristics, as the Spearman Rank-Order Correlation requires that both variables must both be ordinal.

participated under an Awardee Convener (AC) or Designated Awardee Convener (DAC). However, for three clinical episodes, top performers were more likely to participate under an AC or DAC. Few hospitals participated in the 30-day length of episode across clinical episodes, although top performers in two clinical episodes were more likely to have selected 30-day episodes (Exhibit 45).

Across clinical episodes, top performing hospitals shared some similar characteristics. In general, hospitals with higher standardized NPRA were likelier to be smaller, have fewer residents per bed, and have fewer discharges for the clinical episode in 2011. Higher performing hospitals were associated with having fewer hospital beds in 10 out of 14 clinical episodes, and this correlation was statistically significant for two clinical episodes ($p<0.10$). Higher performing hospitals were more likely to have lower resident-to-bed ratios in 11 out of 14 clinical episodes, and this correlation was statistically significant for four clinical episodes ($p<0.10$). Higher performance was correlated with fewer discharges in the clinical episode in 2011 in 10 out of 14 clinical episodes. This correlation was statistically significant for three clinical episodes ($p<0.10$).

In eight of the 14 clinical episodes, top performing hospitals were less likely to be non-profit relative to lower performers (see Exhibit 44). While at least one top performing hospital in each clinical episode had prior bundled payment experience, the proportion with prior experience was never greater than that of average and bottom performing hospitals. The relationship with pay-for-performance is somewhat different. Top performers in five clinical episodes were more likely than average and bottom performers to have prior pay-for-performance experience.

There was some variation with regards to the supply of PAC providers in the market. However there was no clear pattern with respect to performance and the number of SNF beds per 10,000 residents or if the hospital was located in a market with an IRF.

Exhibit 44: Spearman Rank-Order Correlation between Standardized NPRA and Model, Hospital, and Market Characteristics, Model 2 Hospitals

Clinical Episode	Average Number of Selected Clinical Episodes, Q4 2013-Q4 2016 (Rho)	Bed Count, 2011 (Rho)	Resident to Bed Ratio, 2011 (Rho)	Disproportionate Share Percent, 2011 (Rho)	Number of Discharges within Clinical Episode, 2011 (Rho)	Number of SNF Beds/10,000 Residents, 2011 (Rho)
Acute myocardial infarction	**0.42**	0.15	-0.04	0.25	-0.31	-0.15
Cardiac arrhythmia	-0.15	0.12	0.23	**0.37**	0.14	0.17
Cellulitis	0.10	-0.15	-0.17	0.20	**-0.41**	**-0.47**
Congestive heart failure	0.11	**-0.22**	-0.16	-0.12	**-0.20**	-0.08
COPD, bronchitis, asthma	0.05	-0.16	-0.01	-0.06	0.07	**0.19**
Hip and femur procedures except major joint	0.10	-0.04	0.16	-0.16	-0.02	0.01
Medical non-infectious orthopedic	0.15	0.07	-0.28	-0.14	0.10	-0.26
Major joint replacement of the lower extremity	-0.11	0.03	0.06	-0.08	-0.01	0.11
Percutaneous coronary intervention	0.13	-0.08	**-0.37**	-0.10	-0.24	0.00
Renal failure	0.09	-0.14	**-0.36**	-0.04	-0.15	-0.27
Sepsis	0.17	-0.04	-0.13	0.03	-0.03	-0.18
Simple pneumonia and respiratory infections	**0.22**	-0.01	-0.11	**-0.19**	0.03	0.02
Stroke	0.18	-0.21	-0.17	0.03	-0.15	-0.11
Urinary tract infection	0.09	-0.01	-0.19	-0.08	-0.01	0.04

Note: The estimates in this table are results of a Spearman Rank-order Correlation. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. NPRA = net payment reconciliation amount. COPD = chronic obstructive pulmonary disease. SNF = skilled nursing facility.

Source: Lewin analysis of 2011 Medicare claims, 2011 AHRF, CMS' database for all Q4 2013-Q4 2016 BPCI participating hospital EIs, and Awardee-submitted data collected February 2015 through February 2016 for Model 2 EIs participating in BPCI between Q4 2013-Q4 2015.

Exhibit 45: Model, Hospital, and Market Characteristics, by Performance Group, Model 2 Hospitals

Clinical Episode	Performance Group	EIs (N)	EIs under Awardee Convener or Designated Awardee Convener[1], Q4 2016 (%)	EIs with 30-day Length of Episode, Q4 2013-Q4 2016 (%)	Signed up for Gainsharing Activities, Q4 2016[2] (%)	Ownership: Non-profit, 2011 (%)	Bundled Payment Experience[3] Prior to Joining BPCI (%)	Pay for Performance Experience[3] Prior to Joining BPCI (%)	Located in Market with an IRF, 2011 (%)
Acute myocardial infarction	Top Performers	8	50	13	75	63	25	88	100
	Average Performers	18	61	0	89	72	28	72	67
	Bottom Performers	9	89	0	100	89	11	44	67
Cardiac arrhythmia	Top Performers	10	70	0	90	70	40	80	60
	Average Performers	20	75	0	80	70	20	60	85
	Bottom Performers	10	60	0	90	60	50	70	60
Cellulitis	Top Performers	9	56	0	67	56	11	67	100
	Average Performers	18	78	0	94	83	17	56	89
	Bottom Performers	9	89	0	100	78	11	33	56
Congestive heart failure	Top Performers	37	73	3	81	78	16	62	62
	Average Performers	76	54	8	80	80	14	59	70
	Bottom Performers	38	74	5	74	82	19	58	61
COPD, bronchitis, asthma	Top Performers	28	64	0	86	61	18	79	86
	Average Performers	56	75	5	82	80	20	48	79
	Bottom Performers	29	72	0	79	76	15	48	62
Hip and femur procedures except major joint	Top Performers	11	82	0	91	91	18	45	73
	Average Performers	22	50	0	77	73	14	59	64
	Bottom Performers	11	73	9	100	73	27	45	73
Medical non-infectious orthopedic	Top Performers	10	50	0	70	60	10	70	90
	Average Performers	21	76	0	86	57	19	57	67
	Bottom Performers	11	91	0	100	55	36	73	91

Clinical Episode	Performance Group	EIs (N)	EIs under Awardee Convener or Designated Awardee Convener[1], Q4 2016 (%)	EIs with 30-day Length of Episode, Q4 2013-Q4 2016 (%)	Signed up for Gainsharing Activities, Q4 2016[2] (%)	Ownership: Non-profit, 2011 (%)	Bundled Payment Experience[3] Prior to Joining BPCI (%)	Pay for Performance Experience[3] Prior to Joining BPCI (%)	Located in Market with an IRF, 2011 (%)
Major joint replacement of the lower extremity	Top Performers	64	52	20	77	81	14	57	77
	Average Performers	129	63	10	91	76	13	62	74
	Bottom Performers	65	65	9	81	80	17	52	58
Percutaneous coronary intervention	Top Performers	8	63	0	75	38	25	63	100
	Average Performers	16	44	0	75	69	19	63	75
	Bottom Performers	8	63	0	88	63	38	63	75
Renal failure	Top Performers	11	64	0	73	64	27	73	73
	Average Performers	22	68	0	82	64	14	64	86
	Bottom Performers	12	58	8	92	67	33	75	83
Sepsis	Top Performers	24	63	4	71	54	8	50	79
	Average Performers	50	72	4	90	72	10	52	76
	Bottom Performers	25	72	0	84	68	29	42	72
Simple pneumonia and respiratory infections	Top Performers	29	66	0	86	69	14	59	69
	Average Performers	60	75	8	85	75	15	60	78
	Bottom Performers	30	77	0	77	73	25	57	60
Stroke	Top Performers	14	57	0	64	43	7	57	57
	Average Performers	29	45	3	55	76	15	74	79
	Bottom Performers	15	33	0	67	73	20	67	67
Urinary tract infection	Top Performers	13	77	0	85	62	15	54	85
	Average Performers	28	71	4	89	61	25	57	89
	Bottom Performers	14	71	0	79	64	29	50	86

Note: EI = episode initiator. IRF = inpatient rehabilitation facility. COPD = chronic obstructive pulmonary disease.

[1] The percentage of participants signed under an AC or DAC is based on participants' roles in Q4 2016.

[2] Gainsharing activities means that a hospital has indicated it participated in the gainsharing waiver available under the BPCI initiative. Participation in the waiver does not guarantee that a hospital shared monetary gains.

[3] Prior experience measures were analyzed using data from the first reporting period episode initiators were required to report the measures. These measures only include participants that submitted data. Participants that did not submit complete data are excluded from the counts for these measures.

Source: Lewin analysis of 2011 Medicare claims, 2011 AHRF, CMS' database for all Q4 2013 through Q4 2016 BPCI participating hospital EIs, and Awardee-submitted data collected February 2015 through February 2016 for Model 2 EIs participating in BPCI between Q4 2013 through Q4 2015.

Strategies and Opportunities for Realizing Positive NPRA

By design, hospitals with higher standardized NPRA were associated with statistically significant relative decreases in total Medicare payments during the inpatient stay plus 90 days post discharge from baseline to intervention, compared with hospitals with lower standardized NPRA ($p<0.10$) (Exhibit 46). Higher performance was also associated with relative reductions in post-bundle payments,[54] indicating that higher performing hospitals did not achieve relative reductions in total payments by shifting services after the bundle.

Across the 14 clinical episodes, higher performing hospitals displayed similar strategies for realizing NPRA. Our analysis found that, in all 14 clinical episodes, hospitals with higher standardized NPRA were associated with relative reductions in the share of patients discharged to institutional PAC between the baseline and intervention periods compared to hospitals with lower standardized NPRA (statistically significant for 12 clinical episodes, $p<0.10$). Rather than discharging patients to institutional PAC, higher performing hospitals discharged a relatively greater share of patients to home without PAC services in 13 of the 14 clinical episodes. This correlation was statistically significant in eight clinical episodes ($p<0.10$). Higher performing hospitals also discharged a relatively greater share of patients to home health in nine of the 14 clinical episodes, and this correlation was statistically significant for three episodes ($p<0.10$). Furthermore, across all clinical episodes, higher performing hospitals were associated with relative decreases in the rate of 90-day unplanned readmissions from the baseline to intervention, and this correlation was statistically significant for six of the episodes ($p<0.10$) (Exhibit 46).

Circumstances at baseline may have provided more opportunity for hospitals with higher standardized NPRA to reduce payments during BPCI, relative to hospitals with lower standardized NPRA (see **Appendix N** for detailed results). During the baseline period, higher performing hospitals had greater total Medicare payments during the inpatient stay plus 90 days post discharge relative to hospitals with lower standardized NPRA in 13 of the 14 clinical episodes, and this correlation was statistically significant for six of the 13 episodes ($p<0.10$). Additionally, during the baseline period, higher performing hospitals in 12 clinical episodes discharged a greater share of their patients to institutional PAC and a smaller share of patients home without any type of PAC in ten episodes.[55] These initial circumstances might have facilitated reduced use of institutional PAC and reduced total Medicare payments among hospitals with higher standardized NPRA during BPCI.

Higher performance was also correlated with changes in patient mix from baseline to intervention. Hospitals with higher standardized NPRA were associated with relative decreases in the mean overall HCC scores among patients treated in ten clinical episodes from baseline to intervention, and this correlation was statistically significant for six episodes ($p<0.10$). In 11 of the 14 clinical episodes, higher performing hospitals were also associated with relative decreases in the share of

[54] Only 11 of the 14 clinical episodes had sufficient sample size to risk adjust 90-day post-bundle payments. For all 11 episodes, higher performing hospitals were associated with relative decreases in post-bundle payments, and this correlation was statistically significant for eight episodes, $p<0.10$.

[55] The positive correlation at baseline between performance and the percent discharged to institutional PAC was statistically significant for three episodes ($p<0.10$). The negative correlation at baseline between performance and the percent discharged home without HH services was statistically significant for five episodes ($p<0.10$).

patients eligible for both Medicare and Medicaid from baseline to intervention. This correlation was statistically significant for two episodes ($p<0.10$). Additionally, higher performing hospitals in nine clinical episodes were associated with relative decreases in the age of patients from baseline to intervention, and this correlation was statistically significant for two episodes ($p<0.10$). Finally, higher performing major joint replacement of the lower extremity hospitals were associated with a statistically significant relative increase in the share of non-fracture patients, which are elective procedures, from baseline to intervention ($p<0.10$).[56]

[56] This result is not displayed in Exhibit 46.

Exhibit 46: Spearman Rank-Order Correlation between Standardized NPRA and Changes in Payment, Utilization, and Quality Outcomes, Model 2 Hospitals, Q4 2011 – Q4 2016

Clinical Episode	Total Allowed Payment Amount	Total Part A & B 30-day Post-bundle	Unplanned Readmission Rate, 90-day PDP	ED Use, 90-day PDP	Anchor Hospital Stay LOS	Percent Discharged Home without HH	Percent Discharged with HH	Percent Discharged to Institutional PAC	Age	HCC Overall Score	Percent Medicaid Eligible
Acute myocardial infarction											
Cardiac arrhythmia											
Cellulitis											
Congestive heart failure											
COPD, bronchitis, asthma											
Hip and femur procedures except major joint											
Medical non-infectious orthopedic											
Major joint replacement of the lower extremity											
Percutaneous coronary intervention											
Renal failure											
Sepsis											
Simple pneumonia and respiratory infections											
Stroke											
Urinary tract infection											

Note: Dark green shading indicates that standardized NPRA rank positively correlated with an outcome and was statistically significant at 10%. Light green shading indicates that standardized NPRA rank was positively correlated with an outcome, but was not statistically significant. Light orange shading indicates that standardized NPRA rank was negatively correlated with an outcome, but was not statistically significant. Dark orange shading indicates that standardized NPRA rank was negatively correlated with an outcome and was statistically significant at 10%. A blank, unshaded cell indicates that the outcome cannot be presented due to insufficient sample size. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. The total allowed payment amount includes Medicare program payments plus coinsurance and/or copayments for the inpatient stay plus the 90-day post-discharge period. NPRA=net payment reconciliation amount. ED=emergency department. PDP=post-discharge period. LOS=length of stay. HH=home health. PAC=post-acute care. HCC=hierarchical conditions category. COPD=chronic obstructive pulmonary disease.
Source: Lewin analysis of Medicare claims, enrollment, and net payment reconciliation amount data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q4 2016 (intervention period).

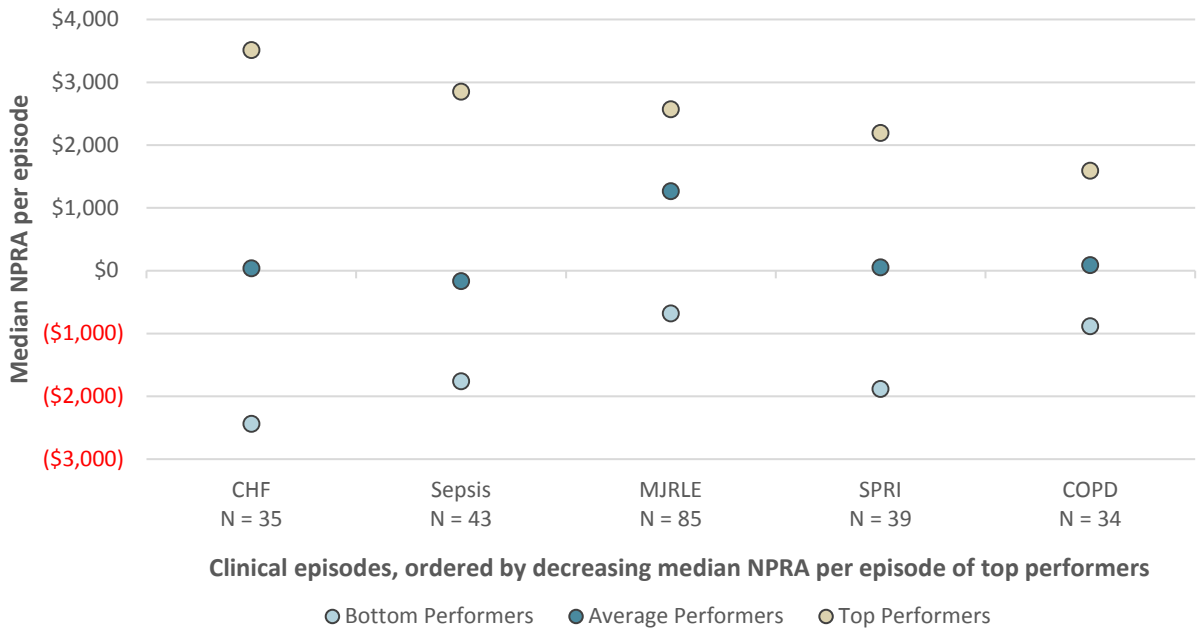
b. BPCI-participating PGPs

Differences in Performance across Participating PGPs

Across the five clinical episodes we analyzed, we found variation in financial performance, as measured by NPRA among BPCI PGPs, as shown in Exhibit 47. Top performers in each clinical episode had a median NPRA that was between $2,472 and $5,952 more than that of bottom performers in the same clinical episode. The episode with the largest difference in median NPRA between top and bottom performers was congestive heart failure. Top performing PGPs participating in this episode had a median NPRA of $3,513, while bottom performers had a median NPRA of -$2,439. Conversely, the episode with the smallest difference in median NPRA between top and bottom performers was COPD, asthma, and bronchitis with a median NPRA of $1,590 among top performers and -$882 among bottom performers.

We also found variation in financial performance across clinical episodes for each performance group. Across the five episodes, the median NPRA for top performers ranged from $1,590 to $3,513. Among average performers, the median NPRA ranged from -$163 to $1,265. Finally, among bottom performers, the median NPRA ranged from -$2,439 to -$682. Compared to other episodes, PGPs participating in congestive heart failure had the largest median gains among top performers ($3,513), while PGPs participating in major joint replacement of the lower extremity had the largest median gains among average performers ($1,265). PGPs in major joint replacement of the lower extremity also had the smallest median loss among bottom performers (-$682), whereas bottom performers in the congestive heart failure episode had the largest median loss (-$2,439).

Exhibit 47: Median NPRA per Episode, by Performance Group, Model 2 PGP, Q4 2013 – Q4 2016



Note: NPRA = net payment reconciliation amount. PGP = physician group practice. CHF = congestive heart failure. MJRLE = major joint replacement of the lower extremity. SPRI = simple pneumonia and respiratory infections. COPD = chronic obstructive pulmonary disease.

Source: Lewin analysis of net payment reconciliation amount data for episodes that began Q4 2013 through Q4 2016.

Correlation between NPRA and Stopping Participation in a Clinical Episode

Across clinical episodes, the financial performance of PGPs is consistent with the proportion of PGP EIs that stopped participating in the clinical episode. Through Q4 2016, top performing PGPs had the lowest dropped episode rate for four out of the five episodes, while bottom performers had the highest dropped episode rate for all five episodes. Across episodes, the dropout rate ranged from 0% to 38% for top performers, 7% to 55% for average performers, and 59% to 100% for bottom performers, as shown in Exhibit 48. Overall, for all performance groups, major joint replacement of the lower extremity had the lowest share of EIs that stopped participating 19%, while COPD, asthma, and bronchitis and simple pneumonia and respiratory infections had the highest at 53%.

Exhibit 48: Episode Initiators that Stopped Participating in the Clinical Episode, by Performance Group, Model 2 PGP, Q4 2013 – Q4 2016

Clinical Episode	Performance Group	EIs (N)	Number that Stopped Participating	Percent that Stopped Participating	Median NPRA per Episode
Congestive heart failure	Top Performers	8	1	13%	$3,513
	Average Performers	18	6	33%	$38
	Bottom Performers	9	8	89%	-$2,439
	Total	35	15	43%	$9
COPD, bronchitis, asthma	Top Performers	8	3	38%	$1,590
	Average Performers	17	6	35%	$89
	Bottom Performers	9	9	100%	-$882
	Total	34	18	53%	$80
Major joint replacement of the lower extremity	Top Performers	21	0	0%	$2,570
	Average Performers	42	3	7%	$1,265
	Bottom Performers	22	13	59%	-$682
	Total	85	16	19%	$1,252
Sepsis	Top Performers	10	2	20%	$2,848
	Average Performers	22	10	45%	-$163
	Bottom Performers	11	7	64%	-$1,759
	Total	43	19	44%	-$180
Simple pneumonia and respiratory infections	Top Performers	10	1	10%	$2,194
	Average Performers	20	11	55%	$53
	Bottom Performers	10	9	90%	-$1,881
	Total	40	21	53%	$41

Note: Median NPRA per episode is representative of all providers within a clinical episode performance group and is not limited to only providers who stopped participation. This table includes PGPs that stopped participating in the clinical episode for any reason, including withdrawing from BPCI. NPRA = net payment reconciliation amount. PGP = physician group practice. COPD = chronic obstructive pulmonary disease.

Source: Lewin analysis of CMS' BPCI database and net payment reconciliation amount data for episodes that began Q4 2013 through Q4 2016.

Differences in Model and PGP Characteristics across Performance Groups

There was little variation among most model characteristics across the five clinical episodes, although we did observe associations between higher performing PGPs and the number of clinical episodes selected and the share of episodes that occurred at a BPCI hospital in the same clinical community (Exhibits 49 and 50).[57,58] Higher performing PGPs in three out of five clinical episodes selected a larger number of clinical episodes to participate in, and this correlation was statistically significant for two clinical episodes ($p<0.05$). For three clinical episodes, higher performing PGPs were more likely to have had at least half of their episodes occur at a BPCI hospital that ever participated in the same clinical community, suggesting that higher performing PGPs may have been able to leverage the knowledge and experience of hospitals with exposure to BPCI.

Across clinical episodes, there was variation in the characteristics of BPCI PGPs and we observed associations between performance and the average NPI count in 2012 and if the PGP was a single-specialty practice. In four out of five clinical episodes, higher performing PGPs were more likely to have fewer providers in their practices, as indicated by the average NPI count in 2012. This correlation was statistically significant for one clinical episode ($p<0.05$). Though overall rates of single-specialty practices were relatively small across all PGPs, PGPs with higher standardized NPRA in major joint replacement of the lower extremity were more likely to be single-specialty practices relative to lower performers.

Exhibit 49: Spearman Rank-Order Correlation between Standardized NPRA and Model and PGP Characteristics, Model 2 PGP

Clinical Episode	Average Number of Selected Clinical Episodes, Q4 2013-Q4 2016 (Rho)	Average NPI Count, 2012 (Rho)	Percent of Nurse Practitioners, 2012 (Rho)[1]	Number of Discharges within Clinical Episode, 2012 (Rho)
Congestive heart failure	0.24	-0.16	0.03	-0.07
COPD, bronchitis, asthma	**0.39**	0.00	-0.14	-0.03
Major joint replacement of the lower extremity	**-0.32**	**-0.27**	**-0.25**	-0.05
Sepsis	**0.37**	-0.13	-0.07	-0.24
Simple pneumonia and respiratory infections	-0.01	-0.12	0.03	-0.23

Note: The estimates in this table are results of a Spearman Rank-order Correlation. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. NPRA = net payment reconciliation amount. PGP=physician group practice. NPI=national provider identifier. COPD=chronic obstructive pulmonary disease.

[1]Clinicians were identified as nurse practitioners according to the Medicare provider specialty associated with their NPI in Part B claims from 2012–2016. We then calculated the percentage of nurse practitioners out of total clinical staff for each PGP in the year 2012.

Source: Lewin analysis of CMS' database for all Q4 2013 through Q4 2016 participating PGP EIs, TIN/NPI crosswalk, Medicare Part A and B claims, and Awardee-submitted data collected February 2015 through February 2016 for Model 2 EIs participating in BPCI between Q4 2013 through Q4 2015.

[57] Clinical communities are a broad classification of clinical episodes and represent the clinical episodes that are most likely to experience exposure to one another in the hospital setting. **Appendix D** contains a list of the clinical episodes included in each clinical community.

[58] We ran a Spearman rank-order correlation for 4 out of 12 characteristics to determine whether higher performance was statistically correlated. We were unable to conduct a Spearman rank-order correlation for the remaining eight binary characteristics, as the Spearman Rank-Order Correlation requires that two variables must both be ordinal.

Exhibit 50: Model and PGP Characteristics, by Performance Group, Model 2 PGP

Clinical Episode	Performance Group	EIs (N)	EI under Awardee Convener or Designated Awardee Convener[1], Q4 2016 (%)	EIs with 30-day Length of Episode, Q4 2013-Q4 2016 (%)	Signed up for Gainsharing Activities[2], Q4 2016 (%)	Intervention Episodes Initiated at a BPCI Hospital in the Same Clinical Community, Q4 2013-Q4 2016 (%)	Single-specialty, 2012-2016 (%)	Hospitalist Groups, 2012-2016 (%)	Prior Bundled Payment Experience[3] Q2 2015-Q2 2016 (%)	Prior Pay for Performance Experience[3] Q2 2015-Q2 2016 (%)
Congestive heart failure	Top Performers	8	100	0	100	25	0	50	0	0
	Average Performers	18	100	0	100	6	6	28	0	0
	Bottom Performers	9	89	11	89	11	11	22	0	11
COPD, bronchitis, asthma	Top Performers	8	100	0	100	13	0	25	0	0
	Average Performers	17	100	0	100	12	6	41	0	0
	Bottom Performers	9	89	11	89	11	11	11	0	11
Major joint replacement of the lower extremity	Top Performers	21	95	10	100	5	52	0	5	14
	Average Performers	42	95	0	100	5	43	0	14	21
	Bottom Performers	22	91	5	100	5	27	5	14	23
Sepsis	Top Performers	10	100	0	100	40	0	40	0	10
	Average Performers	22	100	0	100	14	5	18	0	0
	Bottom Performers	11	100	0	100	9	9	36	0	0
Simple pneumonia and respiratory infections	Top Performers	9	100	0	100	11	0	33	0	0
	Average Performers	20	95	5	100	20	0	30	0	10
	Bottom Performers	10	100	0	100	0	10	30	0	0

Note: PGP=physician group practice. EI=episode initiator. COPD=chronic obstructive pulmonary disease.

[1] The percentage of participants signed under an AC or DAC is based on participants' roles in Q4 2016.

[2] Gainsharing activities means that a PGP has indicated it participated in the gainsharing waiver available under the BPCI initiative. Participation in the waiver does not guarantee that a PGP shared monetary gains.

[3] Prior experience measures were analyzed using data from the first reporting period episode initiators were required to report annual measures. These measures only include participants that submitted data. Participants that did not submit complete data are excluded from the counts for these measures.

Source: Lewin analysis of CMS' database for all Q4 2013 though Q4 2016 participating PGP EIs, TIN/NPI crosswalk, Medicare Part A and B claims, and Awardee-submitted data collected February 2015 through February 2016 for Model 2 EIs participating in BPCI between Q4 2013 through Q4 2015.

Strategies and Opportunities for Realizing Positive NPRA

By design, a Spearman rank-order correlation analysis found that PGPs with higher standardized NPRA were associated with statistically significant relative decreases in 90-day total episode payments from baseline to intervention, relative to PGPs with lower standardized NPRA ($p<0.10$) (Exhibit 51). In three clinical episodes, PGPs with higher standardized NPRA were associated with decreased total payments during the 90-day post bundle period from baseline to intervention (statistically significant in two clinical episodes, $p<0.10$).[59] This suggests that higher performing PGPs did not achieve savings by shifting services after the episode.

Across the five clinical episodes in the PGP analysis, higher performing PGPs displayed similar strategies for realizing NPRA. In all five clinical episodes, higher performance was associated with relative reductions in the share of patients discharged to institutional PAC from baseline to intervention (statistically significant in all five clinical episodes, $p<0.10$). Instead of discharging patients to institutional PAC, in four of the five clinical episodes, higher performing PGPs increased the share of patients discharged home without home health relative to PGPs with lower standardized NPRA (statistically significant in four of the five clinical episodes, $p<0.10$). Higher performing PGPs in four of the five clinical episodes also increased the share of patients discharged to home health relative to lower performing PGPs, but this measure was only statistically significant for one clinical episode ($p<0.10$). Additionally, in four of the five clinical episodes, higher performance was associated with relative decreases in the rate of 90-day unplanned readmissions from the baseline to the intervention period (statistically significant in one of the five clinical episodes, $p<0.10$) (Exhibit 51).

Circumstances during the baseline period may have provided higher performing PGPs more of an opportunity to reduce payments during BPCI, relative to lower performing PGPs. In all five clinical episodes, PGPs with higher standardized NPRA discharged a relatively greater share of patients to institutional PAC in the baseline period, although this was only statistically significant for one of the five clinical episodes ($p<0.10$). In four of the five clinical episodes, higher performing PGPs also discharged a relatively lower share of patients home without home health (statistically significant for one clinical episode, $p<0.10$). For major joint replacement of the lower extremity, higher performance was also associated with higher baseline 90-day readmission rates (statistically significant, $p<0.10$). These initial circumstances might have facilitated reduced use of institutional PAC and reduced total Medicare payments among PGPs with higher standardized NPRA during BPCI.

We also observed that higher performing PGPs in some clinical episodes experienced a change in patient mix from baseline to intervention. In all five clinical episodes, higher performance was associated with relative decreases in the mean HCC overall score among patients treated within the clinical episode from baseline to intervention (statistically significant in two of the five clinical episodes, $p<0.10$). In four of the five clinical episodes, higher performance was also associated with relative decreases in the share of patients eligible for both Medicare and Medicaid from baseline to intervention (statistically significant in one of the five clinical episodes, $p<0.10$). For major joint replacement of the lower extremity, higher performance was associated with relative

[59] We only had a sufficient sample size to risk adjust the post-bundle payments outcome for four of the five PGP clinical episodes.

increases in the share of non-fracture cases and the share of Medicare-Medicaid eligible patients (statistically significant, $p<0.10$).[60]

[60] The correlation between standardized NPRA and the share of non-fracture cases is not displayed in Exhibit 51.

Exhibit 51: Spearman Rank-Order Correlation between Standardized NPRA and Changes in Payment, Utilization, and Quality Outcomes, Model 2 PGP, Q4 2011 – Q4 2016

Clinical Episode	Total Allowed Payment Amount	Total Part A & B 30-day Post-bundle	Unplanned Readmission Rate, 90-day PDP	ED Use, 90-day PDP	Anchor Hospital Stay LOS	Percent Discharged Home without HH	Percent Discharged with HH	Percent Discharged to Institutional PAC	Age	HCC Overall Score	Percent Medicaid Eligible
Congestive heart failure											
COPD, bronchitis, asthma											
Major joint replacement of the lower extremity											
Sepsis											
Simple pneumonia and respiratory infections											

Note: Dark green shading indicates that standardized NPRA rank positively correlated with an outcome and was statistically significant at 10%. Light green shading indicates that standardized NPRA rank was positively correlated with an outcome, but was not statistically significant. Light orange shading indicates that standardized NPRA rank was negatively correlated with an outcome, but was not statistically significant. Dark orange shading indicates that standardized NPRA rank was negatively correlated with an outcome and was statistically significant at 10%. A blank, unshaded cell indicates that the outcome cannot be presented due to insufficient sample size. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. The total allowed payment amount includes Medicare program payments plus coinsurance and/or copayments for the inpatient stay plus the 90-day post-discharge period. NPRA = net payment reconciliation amount. PGP = physician group practice. PDP = post-discharge period. ED = emergency department. LOS = length of stay. HH = home health. PAC = post-acute care. HCC = hierarchical conditions category. COPD = chronic obstructive pulmonary disease.

Source: Lewin analysis of Medicare claims, enrollment, and net payment reconciliation amount data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q4 2016 (intervention period).

E. Medicare Program Savings

As demonstrated in this evaluation, BPCI participants have successfully reduced Medicare payments across a variety of clinical episodes. Under Model 2, 11 of the 32 hospital clinical episodes and 7 of the 21 PGP clinical episodes we examined had statistically significant declines in Medicare allowed payments among BPCI participants relative to the comparison group ($p< 0.10$) and several others had declines in payments that did not reach statistical significance (see Exhibits 10 and 32 in Sections II.B.1 and II.C.1, respectively). However, payment reductions did not translate into net savings to Medicare because they do not account for the reconciliation payments, such as NPRA and other amounts that Medicare paid (e.g., outlier payments) or recovered from participants. To understand the impact of BPCI on Medicare program savings, we estimated the total change in non-standardized payments, that is, the actual provider payments from Medicare,[61] for all of the clinical episodes with sufficient volume by EI type. Then, we computed net savings to Medicare from Q4 2013 to Q4 2016 by subtracting reconciliation payments from the total change in non-standardized payments attributable to BPCI.

CMS retrospectively eliminated NPRA downside risk for some episodes.[62] This meant that participants with payments above the target amount did not have to repay the difference between the target amount and the participants' Medicare episode payments (i.e. NPRA). As a result, reconciliation payments were larger than they would have been if the initiative had been implemented as initially planned. To understand what Medicare savings could have been, we estimated the impact on net savings to Medicare had downside risk not been eliminated.

1. Key Findings

- The DiD estimates of total change in non-standardized payments indicated statistically significant ($p< 0.10$) reductions in payments for episodes initiated by Model 2 hospital EIs ($278.5 million overall, $691 per episode) and for episodes initiated by Model 2 PGP EIs ($255.1 million overall, $726 per episode).
- After accounting for the $735.6 million in reconciliation payments for Model 2 through Q4 2016, BPCI resulted in a statistically significant net loss to Medicare of $202.1 million (-$268 per episode) ($p<0.10$).
- Had CMS not eliminated downside risk, reconciliation payments would have been $389.2 million, and Medicare would have achieved a net savings of $144.3 million for BPCI Model 2 ($191 per episode) through Q4 2016, although this estimate was not statistically significant.

[61] Non-standardized payments vary from the standardized allowed amounts that we use in the DiD analyses. Non-standardized amounts include adjustments for geographic differences in wages and other costs, teaching, and treating a disproportionate share of low-income beneficiaries. These adjustments are removed in the standardized amounts so that DiD analyses can isolate the effect of BPCI.

[62] In November 2014, CMS eliminated downside risk for episodes initiated between Q4 2013 and Q4 2014 due to inaccurate target prices. In July 2016, CMS eliminated downside risk for any episode of care that was initiated as a result of the episode attribution issues caused by incorrect PGP Reassignment Lists initiated in 2015, and in December 2017, CMS extended this elimination of downside risk for all such episodes initiated through Q3 2016.

2. Methods

Net savings to Medicare for BPCI Model 2 from Q4 2013 to Q4 2016 were calculated using the following formula:

> Net savings to Medicare = DiD estimate of total change in non-standardized payments – reconciliation payments

See **Appendix D** for additional details on the definitions and calculations for each component.

3. Results

The results of the DiD analysis showed that episodes initiated by both hospitals and PGPs were associated with statistically significant average declines in standardized payments of $645 and $707 per episode respectively ($p < 0.10$). After multiplying the DiD estimate of change in per-episode standardized payments by the total number of BPCI episodes and adjusting for standardization, the estimates of the total change in non-standardized payments were $278,450,798 for hospital episodes and $255,089,512 for PGP episodes. Combined, there was an estimated decline in non-standardized payments for Model 2 of $533,540,309 (statistically significant at the 10% level).

Reconciliation payments for Model 2 through Q4 2016 were $735,644,161. After subtracting reconciliation payments from the DiD estimate of total change in non-standardized payments, we estimated that BPCI resulted in a net loss to Medicare of $202,103,852 over the first 13 quarters of the initiative (Exhibit 52). This loss was statistically significant ($p<0.10$).

Exhibit 52: Components of Net Savings to Medicare, Model 2, Q4 2013 – Q4 2016

Component	Estimate	LCI	UCI
Total change in non-standardized payments	$533,540,309	$361,164,823	$705,915,796
Reconciliation payments	-$735,644,161	–	–
Net savings to Medicare	-$202,103,852	-$374,479,338	-$29,728,365

Note: The estimates of the total change in non-standardized payments are from difference-in-differences (DiD) models of standardized Medicare paid amounts during the anchor stay and 90 day post-discharge period. The coefficients from DiD estimates of change in per-episode standardized payments are multiplied by -1, multiplied by the number of intervention episodes, and converted to non-standardized amounts. Reconciliation payments include NPRA payments and other payments to BPCI-participating providers due to the initiative. Net savings to Medicare is the difference between the total change in non-standardized payments and reconciliation payments. LCI = lower 90% confidence interval; UCI = upper 90% confidence interval.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers and CMS data on reconciliation payments.

If CMS had not eliminated downside risk, Medicare would have recovered an additional $346,401,616 of NPRA resulting in reconciliation payments of $389,242,245.[63] Subtracting the reconciliation payments (where downside risk was not eliminated) of $389,242,545 from the total change in non-standardized payments results in an estimated net savings to Medicare of $144,297,764 (Exhibit 53). This suggests CMS could have realized a net savings on Model 2 episodes if downside risk had been imposed for some episodes, although this estimate was not statistically significant.

[63] The impact of eliminating downside risk for the first five quarters of the initiative was $38,808,779. The impact of continuing to eliminate downside risk for PGPs until Q3 2016 was $307,592,837.

Exhibit 53: Components of Net Savings to Medicare with NPRA Downside Risk not Eliminated, Model 2, Q4 2013 – Q4 2016

Component	Estimate	LCI	UCI
Total change in non-standardized payments	$533,540,309	$361,164,823	$705,915,796
Reconciliation payments (downside risk not eliminated)	-$389,242,545	–	–
Net savings to Medicare (downside risk not eliminated)	$144,297,764	-$28,077,722	$316,673,251

Note: The estimates of total change in non-standardized payments in this table are from difference-in-differences (DiD) models of standardized Medicare paid amounts during the anchor stay and 90 day post-discharge period. The coefficients from DiD estimates of change in per-episode standardized payments are multiplied by -1, multiplied by the number of intervention episodes, and converted to non-standardized amounts. Reconciliation payments (downside risk not eliminated) include net NPRA payments and other payments to BPCI-participating providers, adjusted for the difference between actual and planned (downside risk not eliminated) NPRA. Net savings to Medicare is the difference between the total change in non-standardized payments and reconciliation payments (downside risk not eliminated).

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers and CMS data on reconciliation payments.

F. Model 2 Discussion

There were 423 hospital and 272 PGP Model 2 EIs during the first 13 quarters of BPCI (Q4 2013 through Q4 2016). Model 2 EIs accounted for 87% of the total episodes initiated under Models 2, 3, and 4 over this period. Hospitals initiated just over half of the episodes under Model 2 and PGPs initiated about 47% of all Model 2 episodes. Both types of EIs participated in all 48 clinical episodes. We were able to conduct impact analyses on 32 hospital clinical episodes and 21 PGP clinical episodes because there was sufficient sample size. Model 2 participants achieved statistically significant relative reductions in total Medicare allowed amounts for 19 out of 53 clinical episode-participant type combinations ($p<0.10$). Even so, Medicare did not achieve savings under Model 2 after accounting for reconciliation payments and other Medicare costs. Claim-based outcomes generally indicated that the quality of care was not affected under BPCI, although patients treated by BPCI hospitals and PGPs reported worse care experiences than comparison patients.

The characteristics of hospitals that chose to participate in BPCI suggest that they may have had more resources to devote to responding to the initiative than those that did not participate. BPCI-participating hospitals were more likely to be larger, non-profit, and urban than non-participating hospitals. They also had higher historical standardized Part A payments than non-participating hospitals. We did not have reliable data on physician affiliation to define non-BPCI PGPs, so we were unable to compare participating PGPs to non-participating PGPs. However, almost all (94%) Model 2 PGP EIs participated under a convener or as a convener, and these collaborations often provided resources that may have supported success under the initiative.

Few Model 2 participants used the waivers from Medicare program requirements and fraud and abuse laws that were offered under BPCI, though many obtained permission to do so in their agreements with CMS. The gainsharing waiver allowed participants to share NPRA or ICS with gainsharing partners, which was a flexibility that was intended to increase engagement and improve collaboration across providers involved in an episode. Approximately two-thirds of Model 2 EIs signed up to use the gainsharing waiver. Among the Awardees that could gainshare, one-third reported distributing 67% of the NPRA available for gainsharing and 38% of the available ICS to their gainsharing partners. Gainsharing partners that received distributions were most often

physicians, typically orthopedic surgeons. Some participants told us that they chose not to participate in gainsharing because their physicians were already highly engaged or because they believed the gainsharing amounts to be insufficient to influence behavior. Fewer participants had gainsharing agreements with PAC providers. Participants we interviewed noted that providing a steady stream of PAC referrals would be a more effective strategy to drive behavior change, and that PAC providers would not be motivated by gainsharing because the disbursements would not compensate for potential losses in revenue if PAC utilization declined.

Hospital and PGP EIs responded to BPCI incentives by reducing Medicare payments, primarily by lowering PAC utilization. For the 32 hospital clinical episodes analyzed, there were relative declines in total Medicare payments for the inpatient stay plus 90 days post-discharge for 24 clinical episodes, 12 of which were statistically significant. For these 12 clinical episodes, BPCI providers reduced total Medicare payments by an average of 4.4% relative to their payments absent BPCI. Among the 21 PGP clinical episodes analyzed, 16 had relative declines in total Medicare payments, seven of which were statistically significant. Relative to payments in the absence of BPCI, PGP EIs reduced total Medicare payments by 6.9% on average across these seven clinical episodes. While both hospital and PGP EIs we interviewed discussed similar strategies to reduce PAC use, there was a pattern of increased HHA payments and reduced SNF and IRF payments across hospital-initiated episodes. Among PGP-initiated episodes, HHA payments tended to decline as well as institutional PAC payments.

Although there were no widespread indications that BPCI had a negative impact on quality of care, we conducted a special analysis of changes in quality for particular subpopulations that might be more vulnerable to reductions in PAC. Specifically, we evaluated the impact of BPCI on the claim- and survey-based quality measures for beneficiaries who were dually eligible for Medicare and Medicaid, were diagnosed with Alzheimer's disease or dementia, or had recently spent time in institutional care. We did not find any indications that the impact of BPCI was systematically different for these vulnerable subpopulations than for the BPCI population as a whole. Further, this analysis confirmed that BPCI participants reduced episode payments even for beneficiaries who were more complex than the average beneficiary in a given clinical episode.

Although beneficiary survey respondents treated by BPCI participants reported less favorable care experiences than the comparison group, self-reported change in functional status from before to after the hospitalization was comparable between the two groups. The differences in care experience were generally small in magnitude, however, the negative association between BPCI and care experience was consistent across clinical episodes and survey waves and for hospital and PGP EIs. Because beneficiary surveys were conducted after the start of the BPCI initiative, we cannot be certain whether these differences were caused by BPCI or factors that existed prior to the initiative.

We evaluated whether EIs were changing to a less resource intensive mix of patients during the intervention period to reduce their episode payments, rather than improving care delivery or coordination. We examined patient characteristics that were associated with resource use to identify any changes in the mix of patients. The average resource intensity of patients did not change relative to the comparison group for the majority of clinical episodes, and when there were differences, they were almost as likely to be increased intensity as decreased intensity. However, for non-fracture MJRLE episodes treated by both hospital and PGP EIs and spinal fusion (non-

cervical) treated by hospital EIs, the BPCI patients were less resource intensive in the intervention period relative to the change for the comparison group. These results are notable because the episodes are for planned procedures, so the EIs may have had opportunities to select the patients they treated.

The variation in financial performance of BPCI participants was related to the changes they enacted under the initiative and their circumstances at baseline. Top performing hospitals and PGPs (i.e., those with the largest average NPRA per episode relative to their target price) discharged a larger share of their patients home instead of to an institutional PAC setting and reduced the unplanned readmission rate. They also had higher baseline institutional PAC use, which may have made it easier for them to reduce PAC payments. While top performing hospitals and PGPs made the largest shifts in PAC use, they also experienced the most favorable changes in patient mix from the baseline to the intervention period.

The reductions in Medicare allowed payments were not an artifact of providers leaving the initiative. Over the first 13 quarters of the initiative, 54% of the matched hospital EIs and 56% of the matched PGP EIs participating in any of the clinical episodes included in this evaluation stopped participating in at least one clinical episode. We conducted an additional analysis based on intent-to-treat methods and found that eight of the 12 hospital clinical episodes and all seven PGP clinical episodes that exhibited a statistically significant relative decline in total payments continued to have a statistically significant decline when we included the episodes of all participants, even after participants withdrew. Therefore, we concluded that the BPCI impact estimates of relative reductions in payments are robust to selective withdrawal for the majority of Model 2 hospital and PGP clinical episodes.

In our evaluation of health care market changes due to BPCI we found a small decline in median market share for BPCI-participating hospitals for six clinical episodes. For PGP EIs that started Q3 2015, there was a small increase in median market share for MJRLE from the baseline to the intervention period. There was no change in the number of HHA or SNF providers to which hospital EIs discharged their patients. Nor was there any change in the number of HHAs or SNFs that received MJRLE patients from PGP EIs. Hospital and PGP EIs (for MJRLE) discharged a larger proportion of beneficiaries to SNFs with high star ratings in the intervention period. Thus we found no relationship between BPCI participation and increased market share or discharges focused on a narrower group of PAC providers. We did find evidence that participants discharged a larger proportion of beneficiaries to SNFs with high star ratings, which suggests that EIs may be encouraging patients to consider quality when selecting PAC.

The statistically significant declines in Medicare allowed payments did not directly translate into Medicare program savings. After accounting for amounts that Medicare paid to or received from participants, we estimated a net loss to Medicare of $202.1 million (-$268 per episode) for BPCI Model 2 ($p<0.10$). BPCI participants did not have to repay amounts in excess of their target amounts for a portion of the intervention period. If negative NPRA had been retrieved from participants as originally planned, we estimated that Medicare would have realized $144.3 million in savings due to BPCI Model 2.

III. Model 3 Results and Discussion

A. Characteristics of the Initiative and Participants

This section summarizes the characteristics of the Model 3 BPCI participants during the first 13 quarters of the initiative, from Q4 2013 through Q4 2016.

1. Key Findings

- 135 Awardees that represented 873 SNF EIs, 144 PGP EIs, 116 HHA EIs, 9 IRF EIs, and one LTCH EI participated in the risk-bearing phase in Model 3 of BPCI. These participants initiated 88,680 episodes of care during the first 13 quarters of the initiative.
- BPCI-participating SNF and HHA EIs were larger and more likely to be for-profit and members of a chain than non-participating providers. They also had higher standardized Part A payments prior to joining BPCI for the clinical episodes they selected.
- Model 3 BPCI participants' use of waivers for Medicare program requirements and fraud and abuse laws was limited except for gainsharing, in which 71% of Awardees that were allowed to use the waiver distributed gainsharing amounts.
- During the first 13 quarters of the initiative, 27% of all Model 3 SNF EIs and 30% of all Model 3 HHA EIs withdrew from BPCI.

2. Methods

The same methods described in Section II.A.2 above were used in the analyses described in this chapter. See **Appendix D** for more information on variable definitions and for details of the qualitative methods.

3. Results

a. Overall Model participant characteristics

There were 135 Awardees that represented 873 SNF EIs, 144 PGP EIs, 116 HHA EIs, 9 IRF EIs, and one LTCH EI in the risk-bearing phase under Model 3 of BPCI. [64] These participants initiated 88,680 episodes of care during the first 13 quarters of the initiative. In the broader context of post-acute care providers, approximately 5% of all SNFs and 1% of all HHAs participated in BPCI.[65]

Of those that participated in Model 3, the average length of participation in BPCI at the end of Q4 2016 was five quarters for SNF EIs, six quarters for HHA EIs, seven quarters for IRF EIs, nine quarters for the LTCH EI, and three quarters for PGPs. The majority of EIs (93% of SNF EIs, 74% of HHA EIs, 89% of IRF EIs, and 99% of PGPs) joined Phase 2 of BPCI in the last two quarters that they were able to enroll, namely Q2 and Q3 2015.

[64] PGPs can participate in Model 2 and Model 3, but cannot simultaneously participate in the same episode in both Models. There were five PGPs that participated only in Model 3. All other PGPs that were in Model 3 also participated in Model 2 clinical episodes. Please see Section II.C for more information about Model 2 PGPs.

[65] These percentages are based on the number of BPCI participating EIs that had an admission for the MS-DRG(s) in which they participated.

Relatively few Model 3 EIs reported experience with care redesign initiatives before participating in BPCI, and even fewer had experience with innovative payment models (see **Appendix E** for the percentage of SNF, HHA, and PGP EIs with prior experience in care redesign and payment incentives). However, Model 3 EIs had experience with ACOs; approximately 30% of BPCI SNF episodes and 26% of BPCI HHA episodes through Q4 2016 included a beneficiary who was aligned with a Medicare accountable care organization (ACO).[66]

The average Model 3 EI participated in 10 clinical episodes; only six EIs participated in all 48 clinical episodes. The most popular clinical episode among Model 3 participants was major joint replacement of the lower extremity (MJRLE), in which 48% of EIs participated (see **Appendix E** for counts by clinical episode). Simple pneumonia and respiratory infections (SPRI) was the second most common clinical episode, chosen by 40% of all Model 3 EIs. Only the combined anterior posterior spinal fusion clinical episode had less than 10% of EIs participating. The majority of Model 3 EIs chose 90-day episodes.

During the first 13 quarters of the initiative, a number of EIs stopped participating in some or all of their BPCI clinical episodes. In Model 3, 496 SNFs (57%), 55 HHAs (47%), and 136 PGPs (94%) stopped participating in at least one clinical episode. Through Q4 2016, 236 SNFs (27%), 35 HHAs (30%), and 103 PGPs (72%) withdrew completely from the initiative.[67]

b. SNF characteristics

Exhibits 54a and 54b compare the BPCI-participating SNF EIs to non-participating SNFs. A higher proportion of BPCI-participating SNF EIs were for-profit organizations (86%) compared with non-participating SNFs (70%) and were more likely to be located in urban areas (84% vs. 70%). Participating SNFs were more likely to be part of a chain (52% vs. 22%), they had higher average bed counts (122 vs. 112), and they averaged more admissions for BPCI episode MS-DRGs in 2011 (136 vs. 94), which was before BPCI was announced, than non-participating SNFs.

[66] Defined as participation in the Medicare Shared Savings Program, Pioneer ACO Model, or Next Generation ACO Model.

[67] Data are limited to SNFs, HHAs, and PGPs that participated in BPCI; these figures do not capture BPCI patients in Models 2 and 4 that are discharged to non-BPCI SNFs and HHAs. These provider counts differ from the number of EIs participating in BPCI because not all EIs had an admission for the MS-DRG(s) in which they participated.

Exhibits 54a and 54b: Baseline Characteristics of BPCI-participating SNF Episode Initiators and Non-participating SNFs, Model 3

Domain	Characteristic	BPCI SNF Episode Initiators (N)	BPCI SNF Episode Initiators (%)	Non-participating SNFs (N)	Non-participating SNFs (%)
Ownership	For Profit	740	86%	9,374	70%
	Government	3	0%	617	5%
	Non-Profit	121	14%	3,311	25%
Urban/Rural	Rural	140	16%	3,938	30%
	Urban	724	84%	9,364	70%
IRF in CBSA	Yes	488	56%	7,238	54%
Hospital-Based	Yes	7	1%	586	4%
Part of Chain	Yes	216	52%	2,946	22%

Characteristic	BPCI SNF Episode Initiators (mean)	Non-participating SNFs (mean)
Bed Count	122	112
Number of Admissions for BPCI Episode MS-DRGs, 2011	136	94
SNF Market Share	6%	6%
Nursing Home Overall Score*	3.45	3.32

* This indicates the number of points out of 5 in overall rating and in three areas: Quality, Survey/Health Inspections, and Staffing. The closer to 5 the better the quality, inspections, and staffing.

Note: Data from 864 BPCI SNF episode initiators and 13,302 non-participating SNFs are included in this exhibit. Non-participating SNFs are all other SNFs not participating in any BPCI initiative and that reported values for all measures listed in the table above. BPCI-participating SNFs that received Medicare certification after 2011 are not included in this table. SNF=skilled nursing facility. IRF=inpatient rehabilitation facility. CBSA=core-based statistical area. MS-DRG=Medicare Severity-Diagnosis Related Group.

Source: Lewin analysis of 2013 Provider of Service (POS) files and 2011 Medicare claims.

In 2011, prior to when BPCI was announced, BPCI-participating SNFs had higher standardized Part A payments during the 90 days following SNF admission for patients admitted with BPCI MS-DRGs than non-participating SNFs (see **Appendix E** for 2011 standardized payments). The difference varied by clinical episode, from 7% higher for stroke admissions to 1% higher for congestive heart failure (CHF) admissions.

c. HHA characteristics

Exhibits 55a and 55b compare the Model 3 BPCI-participating HHAs to all non-participating HHAs. A higher proportion of participating HHAs were part of a chain (73% vs. 32%) and for-profit (81% vs. 76%). BPCI-participating HHAs had more employed nurses on average than did non-participating HHAs (29 vs. 9), although the BPCI average is driven by one large HHA that

had 1,558 nurses. BPCI-participating HHAs also had a greater number of admissions for BPCI episode MS-DRGs during 2011 (374 vs. 101). [68]

Exhibits 55a and 55b: Baseline Characteristics of BPCI-participating HHA Episode Initiators and Non-participating HHAs, Model 3

Domain	Characteristic	BPCI HHA Episode Initiators (N)	BPCI HHA Episode Initiators (%)	BPCI HHA Episode Initiators (N)	BPCI HHA Episode Initiators (%)
Ownership	For Profit	94	81%	7,458	76%
	Government	0	0%	612	6%
	Non-Profit	22	19%	1,699	17%
Urban/Rural	Rural	25	22%	1,886	19%
	Urban	91	78%	7,883	81%
Part of Chain	Yes	85	73%	3,110	32%

Characteristic	BPCI HHA Episode Initiators (mean)	Non-participating HHAs (mean)
Number of Employed Nurses in HHA	29	9
Number of Admissions for BPCI Episode MS-DRGs, 2011	374	101

Note: Data from a total of 116 BPCI HHA episode initiators and 9,769 non-participating SNFs are included in this exhibit. Non-participating HHAs are all other HHAs not participating in any BPCI initiative that reported values for all measures listed in the table above. BPCI-participating HHAs that received Medicare certification after 2011 are not included in this table. HHA=home health agency. Medicare Severity-Diagnosis Related Group

Source: Lewin analysis of 2013 Provider of Service (POS) files and 2011 Medicare claims.

In 2011, BPCI-participating HHAs had higher standardized Part A payments during the 90 days from the start of receiving home health than non-participating HHAs for two of the three clinical episodes examined (see **Appendix E** for 2011 standardized payments). The standardized Part A payments were higher for MJRLE and SPRI admissions in BPCI-participating HHAs (6% and 5% respectively) and virtually the same for CHF admissions.

d. Model incentive structure characteristics

Conveners in BPCI

As described in prior evaluation reports, conveners provide a range of services for their EIs. (See **Appendix A** for the definition of each type of convener.) Approximately 99% of Model 3 EIs participated under or as a convener, which is slightly higher than the 94% of Model 2 EIs. Services offered by conveners in Model 3 were similar to those offered by Model 2 conveners. See Exhibit 4 in the Model 2 participant characteristics section (Section II.A.3) for a summary of the key roles that conveners play in BPCI.

[68] After excluding the large HHA, BPCI participating HHAs employed an average of 16 nurses and had an average of 257 admissions for BPCI episode MS-DRGs during 2011.

Waiver use

The BPCI initiative allows the waiver of Medicare program requirements with respect to telehealth and post-discharge home visits for all Model 3 Awardees to facilitate the implementation of care redesign. Awardees may provide beneficiary incentives or engage in gainsharing agreements under BPCI that may be protected under the BPCI fraud and abuse law waivers, if the Awardee describes plans to use them in its CMS-accepted implementation protocol. About 74% of Model 3 EIs could provide beneficiary incentives because their Awardee had included plans for doing so, but only 17% of those eligible used the waiver (Exhibit 56). With regards to gainsharing, 24 (18%) Awardees entered into gainsharing agreements from Q4 2013 through Q4 2016. Few EIs used the telehealth (2%) or post-discharge home visit waivers (1%).

Exhibit 56: Participation in Various Waivers, Model 3, Q4 2013 – Q4 2016

Waivers	Participants Allowed to Use Waiver (N)	Share of Participants Allowed to Use Waiver of all Participants (%)	Participants that Used Waiver (N)	Share of Participants that Used the waiver of those Allowed to Use Waiver (%)
Telehealth	934	100%	18	2%
Post-Discharge Home Visit	934	100%	5	1%
Beneficiary Incentives	693	74%	119	17%
Gainsharing*	24	18%	17	71%

* The data for the gainsharing waiver are based on Awardees, while the three-day hospital stay waiver, beneficiary incentives, telehealth, and post-discharge home visit are based on episode initiators (EIs). There were 136 Awardees with a CMS-accepted implementation protocol through Q3 2016 included in the gainsharing statistics. The gainsharing waiver use figures are based on Awardee-submitted data through Q2 2017. The data for the other four waivers are based on EIs that initiated at least one BPCI episode from Q4 2013 through Q4 2016.

Note: The 934 EIs that participated and had one BPCI episode between Q4 2013 to Q4 2016 were under 117 Model 3 Awardees. EIs include SNF, HHA, PGP, IRF, and LTCH. The telehealth and post-discharge home visit waivers are available to all Model 3 EIs without specifying it in their Awardee's implementation protocol. Therefore, the denominator used to calculate percent of Model 3 EIs that used these two waivers is the total number of Model 3 EIs through Q4 2016. SNF=skilled nursing facility. HHA=home health agency. PGP=physician group practice. IRF=inpatient rehabilitation facility. LTCH=long term care hospital.

Sources: Lewin analysis of Awardee implementation protocols for Q4 2016 BPCI participants, Medicare claims data for episodes initiated Q4 2013 through Q4 2016, and Awardee-submitted data Q4 2013 through Q4 2016. For gainsharing, Lewin analysis of Awardee implementation protocols for Q3 2016 BPCI participants and Awardee-submitted data Q4 2013 through Q2 2017.

Beneficiary incentives

Although most EIs that could provide beneficiary incentives did not, some interviewees indicated that they were providing or considering beneficiary incentives, such as personal emergency response systems or equipment (e.g., blood pressure cuff, weight scale, pulse oximeter). Equipment was the most widely distributed incentive, although it was provided by only five Awardees (Exhibit 57).

Exhibit 57: Beneficiary Incentives Distributed by Episode Initiators to Beneficiaries, Model 3, Q4 2013 – Q4 2016

Incentive Description	Awardees Allowed to Use Incentives	EIs Allowed to Use Incentives and Had at Least One BPCI episode	Awardees that Provided One or More Incentives	Beneficiaries Receiving One or More Incentives	Average Cost per Incentive Provided
Transportation	18	39	2	14	$181
Equipment	9	627	5	374	$32
Home care/home visits	6	453	2	52	$351
Living arrangement services	3	567	3	52	$228
Telehealth/technology	27	638	10	93	$78
Wellness program/resources	3	576	0	0	
Medication management tools	2	56	1	1	$170

Note: A blank value indicates we could not calculate the average cost per incentive due to missing data. EI=episode initiator.

Source: Lewin Program Adherence Report based on analysis of Awardee-submitted data regarding distributed beneficiary incentives through Q4 2016.

Telehealth and post-discharge home visit waivers

Similar to findings in previous annual reports, telehealth and post-discharge home visit waivers were not mentioned often during interviews with Model 3 participants. One interviewee planned to pilot test telehealth hardware and software with a referral partner to be able to use the telehealth waiver. A Medical Director at another site developed a separate company to provide care to patients in the home setting under the post-discharge home visit waiver.

Gainsharing

Among the 24 (18%) Model 3 Awardees that had plans to gainshare, 17 distributed $19.8 million in net payment reconciliation amounts (NPRA) to their gainsharing partners between Q4 2013 and Q2 2017, which equals about 21% of the total NPRA available for gainsharing through Q3 2016. No Model 3 Awardees distributed ICS.

Institutional PAC providers were the most common type of gainsharing partners (Exhibit 58). On average, each institutional PAC partner received $24,989 in NPRA from Q4 2013 through Q2 2017. Physicians and PGPs were also common types of gainsharing partners and received average distributions of $54,472 and $31,974, respectively. The majority of physicians to receive a gainsharing distribution were orthopedic surgeons. On average, each gainsharing orthopedic surgeon received $71,893 in NPRA.

Exhibit 58: Gainsharing Distributions Received by Partner Type, Model 3, Q4 2013 – Q2 2017

Gainsharing Partners	Awardees that Reported Eligible Partners of This Type	Number of Partners Receiving a Distribution	Number of Partners Receiving NPRA	Number of Partners Receiving ICS
Physicians	12	95	95	0
PGPs	9	40	40	0
Hospitals	8	21	21	0
Institutional PAC	12	309	309	0
HHAs	8	20	20	0
Other	2	18	18	0

Note: NPRA=net payment reconciliation amount. ICS=internal cost savings. PGP=physician group practice. PAC=post-acute care. HHA=home health agency.
Source: Lewin analysis of Awardee-submitted data collected February 2016 through August 2017 for Model 3 Awardees participating in BPCI from Q4 2013 through Q2 2017.

During interviews and site visits, we gathered Model 3 participant views on gainsharing. Representatives of organizations participating in gainsharing indicated that they were most likely to be involved in gainsharing agreements with a parent entity (e.g., corporate, system, or convener). Few interviewees mentioned gainsharing with hospitals, physicians, or downstream PAC providers. Interviewees that did gainshare with other organizations mentioned doing so to obtain buy-in, increase engagement, or incentivize certain behaviors, including promoting collaboration and achieving quality targets, such as readmissions. Challenges included a lack of hospital interest in gainsharing, partner participation in other financial incentives (e.g., co-management) that were more lucrative than gainsharing, and the 50% cap on gainsharing limiting the interest of physicians.[69] A handful of interviewees noted they were not participating in gainsharing and described their rationale. One of the interviewees mentioned that their ability to reduce acute care readmissions was more attractive to their hospital partners than gainsharing. Another interviewee, not participating in gainsharing, stated that the 50% cap on physician gainsharing payments, lag in payments, and administrative requirements of the waiver made participation less attractive to partners.

B. Impact of BPCI among Participating SNFs

1. Payment, Utilization, Quality, and Patient Mix

This section presents the BPCI impact estimates on payments, utilization, quality, and characteristics of the mix of patient for episodes initiated at skilled nursing facilities in the first 13 quarters of the initiative, Q4 2013 through Q4 2016.

[69] Gainsharing amounts are subject to a cap set at 50% of the total Medicare FFS expenditures included in the episodes attributed to the physician.

a. Key Findings

- Among the 11 SNF-initiated clinical episodes with sufficient sample size, there were eight clinical episodes with a reduction in the total payment amount included in the bundle definition for 90-day episodes, and the decrease was statistically significant for seven clinical episodes.
- The reduced payments were due to the relative decline in SNF standardized allowed payments in the 90-day PDP from the qualifying hospital stay because of fewer days in the SNF. Eight of the 11 clinical episodes had statistically significant declines in SNF payments and in the number of SNF days relative to the comparison group.
- HHA payments increased for nine of the 11 clinical episodes relative to the change in the comparison group and the increase was statistically significant for five clinical episodes.
- Across the 12 Model 3 SNF clinical episode strata, five had indications that BPCI patients were less resource intensive in the intervention period relative to the comparison group. One stratum appeared to have a shift towards a more resource intensive patient mix, and six strata had no consistent pattern of changes in resource intensity.

b. Methods

The same methods described in Section II.B.1 above were used in the analyses described in this section. We constructed comparison groups for 11 Model 3 SNF clinical episodes that were deemed to have a sufficient sample size for meaningful analysis. The episodes in these 11 clinical episodes represent 73% of all episodes initiated by Model 3 SNF EIs during the first 13 quarters of the initiative. See **Appendix D** for a more detailed description of the methodology.

c. Results

Sample characteristics

We describe the BPCI Model 3 SNF analytical sample for the clinical episodes for which we had sufficient sample size in Exhibit 59. The number of BPCI SNFs included in the sample ranged from 78 to 236, and the number of episodes ranged from 727 to 6,615 episodes for a given clinical episode over the first thirteen quarters of the initiative. Because providers could join BPCI over an extended period and could terminate participation in a clinical episode at any given quarter or withdraw from the initiative completely, providers in this analysis participated on average six quarters from Q4 2013 to Q4 2016. During this period, 236 of 873 SNF EIs (26%) withdrew entirely from the initiative. Among the 493 SNFs participating in any of the 11 clinical episodes analyzed in this report, approximately 27% stopped participating in at least one clinical episode during the first thirteen quarters. The BPCI impact estimates include episodes from all hospital EIs during their participation period. The contribution of episodes from hospital EIs that stopped participating by Q4 2016 varies by clinical episode. Across the 11 Model 3 SNF clinical episodes, 34% of the intervention episodes were initiated by SNFs that later terminated their participation in a clinical episode by the end of the Q4 2016. Across clinical episodes, there was a large difference in the percentage of EIs that discontinued their participation, with as few as 9% of EIs in hip and femur procedures except major joint and as many as 64% of EIs participating in renal failure.

Exhibit 59: Characteristics of the Matched BPCI Providers Included in the BPCI Impact Estimates, Model 3 SNF, Q4 2013 – Q4 2016

Clinical Episode	BPCI SNFs (#)	Matched Intervention Episodes (#)	Average Length of Participation (Quarters)	EIs that Terminated Participation in the Clinical Episode (#)	Episodes from EIs that Terminated (%)
Congestive heart failure	181	2,977	6	59	45.3
COPD, bronchitis, asthma	98	903	6	42	62.7
Hip and femur procedures except major joint	119	2,440	6	12	9.1
Major joint replacement of the lower extremity	218	6,615	6	28	10.5
Medical non-infectious orthopedic	125	2,247	6	44	50.1
Other respiratory	78	727	6	32	60.8
Renal failure	97	1,421	6	37	63.8
Sepsis	194	4,471	6	60	46.1
Simple pneumonia and respiratory infections	236	2,919	6	63	39.4
Stroke	98	1,365	6	33	46.1
Urinary tract infection	153	2,036	6	62	53.4

Note: The sample does not include BPCI providers for which no comparison provider was found; 30% to 64% of BPCI SNFs were matched to comparison SNFs. This exhibit is limited to the matched BPCI providers used to calculate the DiD results in the remainder of this section. See **Appendix D** for information on the methods used to determine the sample. DiD=difference-in-differences. SNF=skilled nursing facility. EI=episode initiator. COPD=chronic obstructive pulmonary disease.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2013 through Q4 2016 for BPCI providers.

Has patient mix changed under BPCI?

Because target prices are not risk adjusted, episode payments for BPCI participants could decline relative to their target price if their mix of patients during the intervention period is less resource intensive than their mix of patients during the baseline. Relative to Model 2, under Model 3, there may be more opportunities for participants to change their mix of patients in response to BPCI because PAC providers choose who they will treat. In this section, we examine changes in patient characteristics that are associated with higher resource use across Model 3 SNF strata (including MJRLE stratified into fractures and non-fractures) to assess if the resource intensity of the average patient changed under the initiative. We compared the change between the baseline and intervention period for BPCI patients relative to the change in the matched comparison group of patients in basic demographic characteristics, count of HCCs, utilization of care in the six months prior to the qualifying hospitalization, and diagnostic and functional information from the initial patient assessment conducted at their episode-initiating SNF. For each of the measures in Exhibits 60 and 61, a negative value indicates a decline in the resource intensity of the BPCI patients during the intervention from the baseline period relative to the comparison group. Similarly, a positive value suggests a relative increase in patient resource intensity. (Please note, the impact analysis on payment, utilization, and quality presented above controlled for changes in the claim-based patient characteristics.)

We categorized Model 3 strata into three broad groups: decline in patient resource intensity, increase in patient resource intensity, and no change.[70] Our categorization was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates. (See **Appendix D** for additional details of the categorization.)

Five of the 12 Model 3 strata had indications that BPCI patients were less resource intensive in the intervention period relative to the comparison group. Other respiratory appeared to have a shift towards a more resource intensive patient mix, and six strata had no change in resource intensity.

Among episodes triggered by a SNF stay, comparisons between the patient mix measures based on claims and assessment data demonstrate the differences in information available. For example, the claims data indicated no changes toward a less intensive patient mix, but the addition of assessment measures suggests a less intensive patient mix for five strata. This finding indicates particular challenges with risk-adjusting Model 3 target prices using only claim-based measures.

The four assessment-based functional status measures (moving in bed, transferring, walking in room, and toileting) indicate that BPCI-participating SNFs may have been treating patients who required less assistance after joining BPCI relative to the change for the comparison group. Eight of the 12 strata had a decline in the proportion of patients who needed assistance in all functional status outcomes relative to the comparison group. These declines were statistically significant for three or more measures in three of the eight strata ($p<0.10$). These functional measures, however, may be more subjective than measures indicating the presence of comorbidities, so these results should be interpreted with caution. Additional information on which characteristics had a

[70] The "no change" category includes strata that do not exhibit a consistent pattern toward a decline or an increase in patient resource intensity. This could be because they have indications of both decreases and increases in patient resource intensity or no indications of changes in either direction.

statistically significant change can be found in Exhibits 60 and 61. Definitions of the Minimum Data Set (MDS) characteristics can be found in **Appendix D**.

Exhibit 60: Relative Changes in Claim-based Characteristics, by Clinical Episode Strata, Baseline to Intervention, Model 3 SNF, Q4 2013 – Q4 2016

Change in Patient Mix	Clinical Episode Strata	Number of Episodes Q4 2013- Q4 2016	Age: 80+ years	Medicaid Eligibility	Disabled, no ESRD	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
Less resource intensive	Congestive heart failure	2,977	-1.8	0.8	**2.2**	0.04	-1.0	-1.5	1.0	-3.1
	Hip and femur procedures except major joint	2,440	-0.2	2.5	0.0	-0.06	-1.5	-1.7	-1.7	-1.7
	Major joint replacement of the lower extremity - Fractures	1,804	-3.9	-0.7	0.3	-0.09	-2.1	-1.7	-3.2	-4.5
	Medical non-infectious orthopedic	2,247	-4.0	**4.5**	0.6	0.09	2.8	1.2	-2.9	1.4
	Stroke	1,365	-1.0	3.5	1.6	-0.05	0.5	3.1	-0.9	-1.8
More resource intensive	Other respiratory	727	-7.4	2.2	**6.2**	0.38	**7.7**	0.9	**7.6**	2.8
No consistent pattern	COPD, bronchitis, asthma	9.3	1.8	1.2	1.2	0.12	0.2	1.4	-1.7	3.2
	Major joint replacement of the lower extremity - Non-fractures	4,811	-2.2	**-2.5**	-1.4	0.00	-0.3	-0.1	1.3	-1.8
	Renal failure	1,421	1.2	-0.9	1.5	-0.23	0.7	-2.9	-1.9	-4.0
	Sepsis	4,471	-0.1	2.4	0.0	-0.02	0.8	-0.6	2.0	-1.4
	Simple Pneumonia and respiratory infections	2,919	-0.5	0.1	1.2	0.07	1.1	**-4.7**	-1.0	-0.3
	Urinary tract infection	2,036	-4.1	**7.2**	**2.4**	-0.02	0.2	-1.4	-2.7	-1.3

* These characteristics measure utilization of care in the six months prior to the anchor hospitalization. Count of HCCs is based on the six months prior to the anchor hospitalization.

Note: Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Categorization of resource intensity was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates. See Appendix D for additional details of the categorization. SNF=skilled nursing facility. ESRD=end-stage renal disease. HCC=hierarchical conditions categories. COPD=chronic obstructive pulmonary disease.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q4 2016 (intervention period) for BPCI EIs and the matched comparison providers.

Exhibit 61: Relative Changes of Assessment-based Characteristics of BPCI and Comparison Beneficiaries, by Clinical Episode Strata, Baseline to Intervention, Model 3 SNF, Q4 2013 – Q4 2016

Change in Patient Mix	Clinical Episode Strata	Number of Episodes Q4 2013- Q4 2016	Not Currently Married	Moderate to Severe Cognitive Impairment	Moderate to Severe Depression	Rejected Necessary Evaluation or Care at Least Once	Need Extensive Assistance or are Totally Dependent Moving in Bed	Need Extensive Assistance or are Totally Dependent Transferring	Need Extensive Assistance or are Totally Dependent Walking in Room	Need Extensive Assistance or are Totally Dependent Using the Toilet	Unhealed Pressure Ulcer	Incontinence	Active Diagnosis of Alzheimer's	Active Diagnosis of Dementia	Shortness of Breath	Require Special Treatment[1]
Less resource intensive	Congestive heart failure	2,977	0.5	2.4	-0.9	-1.3	-4.4	-3.6	-4.4	-3.5	-1.3	-1.4	0.7	1.1	-5.7	-0.9
	Hip and femur procedures except major joint	2,440	1.1	-2.7	-1.6	-0.4	-0.9	-1.1	-0.5	-1.9	-0.8	-6.0	-2.9	0.0	-3.4	-1.3
	Major joint replacement of the lower extremity - Fractures	1,804	-5.0	-5.9	-0.7	-2.1	-4.0	-4.2	-7.8	-3.6	-0.3	-8.0	-1.6	-5.8	-3.0	-0.2
	Medical non-infectious orthopedic	2,247	1.2	-0.8	-0.5	0.2	-4.1	-6.1	-7.5	-5.8	-1.3	-5.5	-0.4	1.5	-5.6	-2.2
	Stroke	1,365	-4.4	-1.9	-0.8	-0.1	-0.7	-2.5	-4.1	-3.1	-0.8	-3.8	0.1	-6.9	-2.6	-3.8
More resource intensive	Other respiratory	727	-2.6	-3.5	-1.4	-0.7	2.5	3.9	0.4	2.0	-1.7	-7.5	-1.2	1.8	9.6	5.3
No consistent pattern	Chronic obstructive pulmonary disease, bronchitis, asthma	903	9.9	-3.5	-3.7	0.5	-3.1	-2.6	-2.7	-0.8	-2.2	4.1	2.4	-0.6	4.1	1.7
	Major joint replacement of the lower extremity - Non-fractures	4,811	-2.1	-0.4	-1.1	0.1	2.5	0.4	1.7	1.2	-0.1	-0.1	0.0	-0.6	-0.9	-1.7
	Renal failure	1,421	0.2	0.2	2.5	0.4	-3.6	-2.4	-1.8	-2.1	-3.0	1.8	-0.9	-3.3	0.0	-6.0
	Sepsis	4,471	1.6	1.3	2.1	2.1	-1.0	-0.3	-2.6	0.5	-2.8	-0.2	0.1	1.5	-0.6	-0.3
	Simple Pneumonia and respiratory infections	2,919	-1.6	2.9	-1.3	-0.3	-2.6	-1.8	-1.4	-0.3	-0.5	-0.2	0.7	3.6	-0.4	1.5
	Urinary tract infection	2,036	1.4	4.6	2.7	1.7	0.9	-0.7	-0.6	0.6	3.5	1.4	-2.3	5.9	-2.8	0.0

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. The Minimum Data Set (MDS) assessment is administered within five days (plus 3 days grace) of the start of care date. The measurement period for outcomes listed above is the seven-day "look-back" period preceding the assessment (the first week of the SNF stay). Assessment data was not available for all episodes. This table is limited to the episodes where we had the initial patient assessment data. Categorization of

resource intensity was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates. See Appendix D for additional details of the categorization. HHA = home health agency. ADL = Activities of Daily Living.
Source: Lewin analysis of Medicare claims, enrollment, and MDS data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q4 2016 (intervention period) for BPCI EIs and the matched comparison providers.

How have average standardized payments changed under BPCI?

Since the implementation of BPCI in Q4 2013 through Q4 2016, the total payment amount included in the bundle definition (Medicare program payments plus coinsurance and/or copayments) declined for the majority of the BPCI SNF-initiated clinical episodes relative to the change in comparison episodes. For 90-day episodes, the total amount included in the bundle decreased from baseline to intervention for BPCI SNFs relative to comparison SNFs in eight of the 11 clinical episodes, and the decline was statistically significant for seven ($p<0.10$) (Exhibit 62). Across these seven clinical episodes, BPCI providers reduced the total payment amount included in the bundle by 5.1% on average relative to their payments absent BPCI (Exhibit 63).[71] Relative reductions in the SNF standardized allowed payments during the 90-day PDP likely contributed to the decline in the total amount included in the bundle. The SNF payments declined in all 11 clinical episodes, and the decline was statistically significant in eight relative to the comparison group ($p<0.05$). Standardized allowed payments for HHA and readmissions increased relative to the comparison group. The HHA payments amount increased in nine of 11 clinical episodes relative to the change in the comparison group, and the increase was statistically significant for five ($p<0.10$). The readmission payments increased in seven of 11 clinical episodes. The increase was statistically significant for two clinical episodes ($p<0.05$), and these were two of the three clinical episodes with a relative increase in the total amount included in the bundle. In addition, one clinical episode had a statistically significant decline in readmission payments. Detailed results of the BPCI impact estimates by clinical episode are located in **Appendix O**.

[71] BPCI providers' payments absent BPCI are calculated as what their payments would have been if they had experienced the same change in payments from the baseline to intervention period as the comparison group.

Exhibit 62: Impact of BPCI on Payment Outcomes, by Clinical Episode, Model 3 SNF, Baseline to Intervention, Q4 2013 – Q4 2016

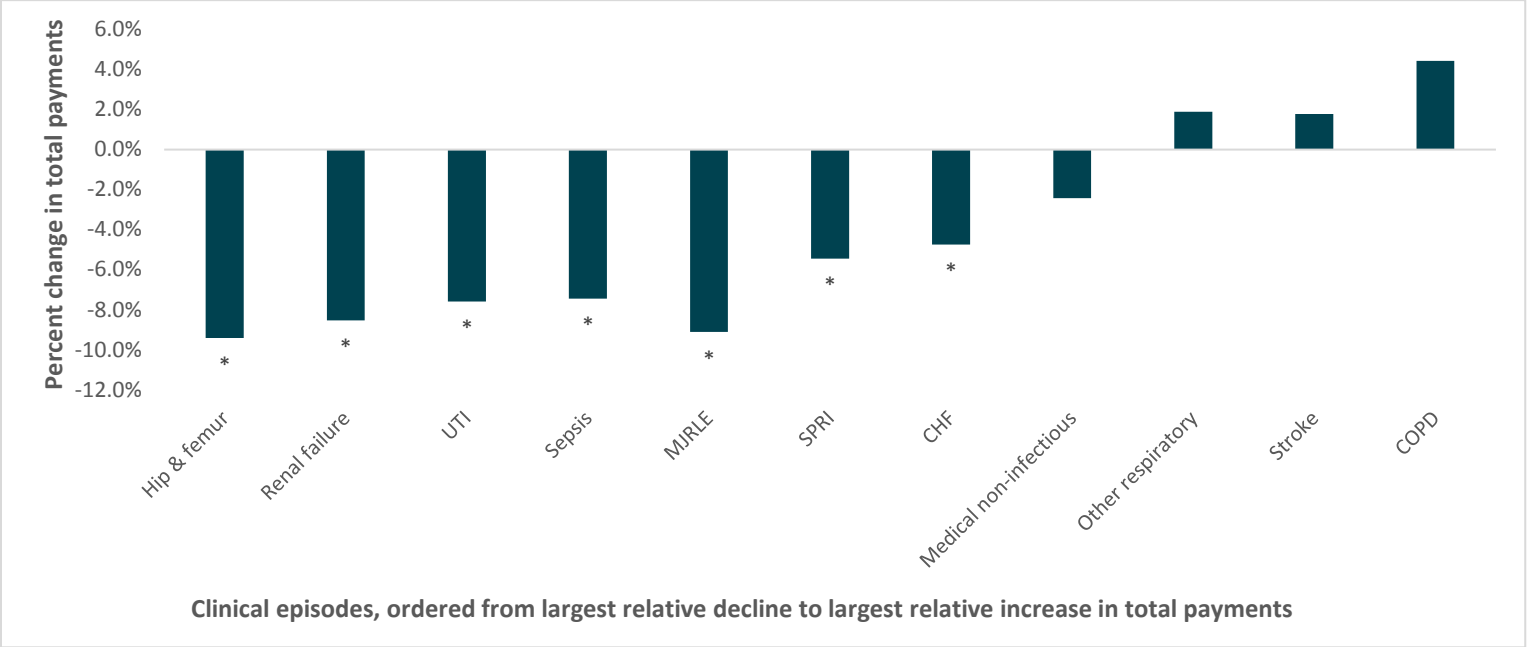
Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Total Amount Included in Bundle Definition[1]	Total Amount Paid by Medicare, IP through 90-day PDP[2]	SNF Standardized Allowed Amount, 90-day PDP[3]	HHA Standardized Allowed Amount, 90-day PDP[3]	Readmissions Standardized Allowed Amount, 90-day PDP[3]	Part B Standardized Allowed Amount, 90-day PDP	Total Part A & B 30-day Post-bundle Period
Congestive heart failure	2,977	-$1,436	-$1,079	**-$1,656**	$179	$145	-$8	-$333
COPD, bronchitis, asthma	903	$1,217	$1,000	-$95	$59	$569	$481	$594
Hip and femur procedures except major joint	2,440	**-$2,991**	-$1,244	**-$3,667**	**$258**	$446	**$263**	-$165
Major joint replacement of the lower extremity	6,615	**-$1,849**	**-$1,412**	**-$1,828**	**$290**	-$107	$7	-$179
Medical non-infectious orthopedic	2,247	-$749	-$870	**-$2,146**	**$210**	$322	$171	-$303
Other respiratory	727	$554	$449	**-$2,685**	$208	**$2,094**	$364	$307
Renal failure	1,421	**-$2,792***	-$853*	-$1,134	$69	-$503	-$45	-$142
Sepsis	4,471	**-$2,270***	**-$2,228***	**-$1,575**	**$204**	-$737	-$72	**-$721**
Simple pneumonia and respiratory infections	2,919	-$1,487	**-$1,439**	**-$1,632**	-$2	-$35	-$2	**-$743**
Stroke	1,365	$574*	$1,951	-$1,369	$36	**$1,172**	**$665**	$233
Urinary tract infection	2,036	**-$2,313**	**-$1,557**	**-$1,999**	-$69	$46	$182	$239

[1]The total amount included in bundle definition values are based on only the 90-day episodes.
[2]Total amount paid does not include beneficiary out-of-pocket expenses.
[3]These payments are not conditional upon use of the service.

* This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate. Equal trends test was conducted for total amount paid by Medicare IP through 90-day PDP, total amount included in the bundle definition for 90-day episodes, emergency department visits, readmission, and mortality outcomes.

Note: The estimates in this table are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented due to insufficient sample size. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. SNF=skilled nursing facility. IP=inpatient. PDP=post-discharge period. HHA=home health agency. COPD=chronic obstructive pulmonary disease.
Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

Exhibit 63: Percent Change in the BPCI Total Payment Amount included in the Bundle Definition from What Payments Would have Been Absent BPCI, by Clinical Episode, Model 3 SNF, Q4 2013 – Q4 2016



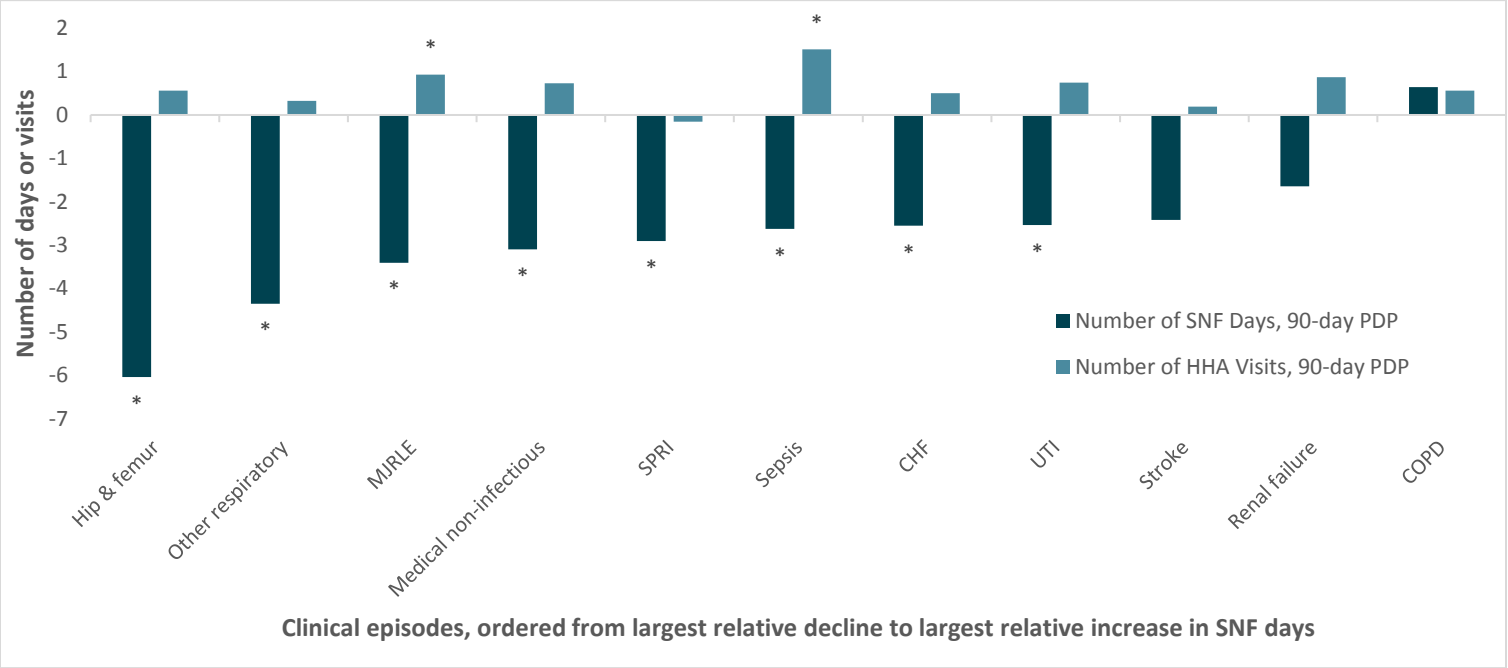
Note: The payments in this exhibit are the risk-adjusted total amount included in the bundle definition for 90-day episodes ordered starting with the largest relative decline. Payments absent BPCI, or the counterfactual, is the BPCI baseline payment amount plus the change in episode payment amount for the comparison group. The counterfactual can be expressed as: BPCI before + (Comparison after – Comparison before). The percent change can then be expressed as: (BPCI after – Counterfactual) / (Counterfactual). Results are sorted by the total Medicare allowed payment amount DiD estimate. SNF=skilled nursing facility. UTI=urinary tract infection. MJRLE=major joint replacement of the lower extremity. SPRI=simple pneumonia and respiratory infections. CHF=congestive heart failure. COPD=chronic obstructive pulmonary disease.

*Indicates DiD estimates are statistically significant at the 10% level.
Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

How have the services changed under BPCI?

The changes in PAC service use were consistent with changes in payments. We observed relative declines in the number of SNF days in the 90-day PDP for 10 clinical episodes and the decline was statistically significant in eight ($p<0.05$) (Exhibit 64). HHA visits increased relative to the comparison group in 10 of the 11 clinical episodes. The two clinical episodes that had statistically significant increases in HHA visits ($p<0.10$), major joint replacement of the lower extremity and sepsis, also had a statistically significant decline in the number of SNF days ($p<0.05$). Detailed results of the BPCI impact estimates by clinical episode are located in **Appendix O**.

Exhibit 64: Impact of BPCI on SNF Days and HHA Visits, by Clinical Episode, Model 3 SNF, Baseline to Intervention, Q4 2013 – Q4 2016



*Indicates DiD estimates are statistically significant at the 10% level.

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. These utilization measures are conditional upon use of the service. Beneficiaries must have spent a minimum of one day in a skilled nursing facility (SNF) and had at least one home health agency (HHA) visit during the 90-day post-discharge period (PDP) to be included in the DiD estimate for number of SNF days and HHA visits, respectively. SNF=skilled nursing facility. HHA=home health agency. UTI=urinary tract infection. MJRLE=major joint replacement of the lower extremity. SPRI=simple pneumonia and respiratory infections. CHF=congestive heart failure. COPD=chronic obstructive pulmonary disease.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

How has quality of care changed under BPCI?

Claim-based results suggest that the quality of care generally did not change under BPCI Model 3 for SNF-initiated episodes (Exhibit 65). For most of the clinical episodes, there were no statistically significant changes for BPCI relative to the comparison group in mortality rates, emergency department use, readmission rates 90 days from the episode start date, or the share of patients that were discharged successfully to the community.[72] There was a statistically significant decrease in the mortality rate for major joint replacement of the lower extremity ($p<0.05$). The unplanned readmission rate increased for stroke ($p<0.10$), consistent with the statistically significant increase in readmissions payments ($p<0.05$). There was a statistically significant decrease in the readmission rate relative to the change in the comparison group for sepsis episodes ($p<0.10$). There was a statistically significant decline in the share of episodes successfully discharged to the community for stroke episodes treated by BPCI SNFs relative to SNFs in the comparison group ($p<0.10$). The sensitivity test results suggest that the statistically significant changes in the BPCI impact estimates for mortality, emergency department use and readmissions may have been a result of the random comparison episode selection, and they may not have been statistically significant if a different sample of comparison episodes had been selected. (Results of the sensitivity test are presented with the BPCI impact estimates in **Appendix H**.) We used patient assessment data from the Minimum Data Set (MDS), which provides information on functional capabilities of SNF users, to conduct additional sensitivity tests about whether the statistically significant changes in quality could be due to changes in patient mix. We added these variables to the risk-adjustment regression models and found that the impact estimates did not generally change. (The variables added to the model are the same ones included in the patient mix tables in Exhibit 61 above).

We also used assessments from the MDS to measure changes in patients' functional status from the beginning to the end of the PAC stay. The assessment-based quality outcomes generally indicated that BPCI episodes had lower rates of improvement in functional status during the intervention than the baseline period relative to the comparison group among the approximately 80% of patients for whom we were able to measure these outcomes.[73] The majority of the outcomes (69%) had declines in the rate of improvement for BPCI episodes relative to episodes from comparison SNFs, and the declines were statistically significant across two clinical episodes, major joint replacement of the lower extremity and hip and femur procedures except major joint ($p<0.05$). There was a statistically significant increase in the proportion of patients with improvements in only two clinical episodes ($p<0.10$).

We conducted additional analyses to better understand the relative declines in functional status and quality. For BPCI episodes, the number of days in the SNF declined, thus reducing the time to improve in functional status during their SNF stay. To investigate whether the declines in

[72] The share of patients who were discharged successfully to the community measures the share of patients who were discharged from the episode-initiating SNF stay to the community, with or without home health services, and remained in the community for 30 days. To remain in the community for 30 days means the patient was not be readmitted to a SNF or inpatient hospital and did not die.

[73] The most common reason we could not measure this outcome was because the patient's SNF stay was not long enough for them to have two assessments. Patients discharged within the first week of their SNF stay only have one assessment because their initial assessment (the five-day PPS assessment) also serves as their discharge assessment.

improvement for BPCI episodes were due to shorter SNF stays, we included the number of days between assessments in the risk-adjustment models. We found that shorter SNF stays may have contributed to the reductions in the rates of improvement, but that does not fully explain the relative decline. [74] Furthermore, because the assessment outcomes measure change during the first SNF stay, we do not know the impact of BPCI on functional status throughout the entire episode of care.

One potential reason for the relative declines in successful discharges of stroke patients could be if those going to a SNF had become more complex. To determine whether the declines in quality for Model 3 SNF stroke episodes were the result of an increase in more complex cases for BPCI SNFs relative to comparison SNFs, we examined the share of episodes discharged from a qualifying hospital that was certified as a Comprehensive Stroke Center, which usually treats more complex patients. We found that while a slightly larger share of BPCI episodes were discharged from certified Comprehensive Stroke Centers, the shares did not change from baseline to intervention for BPCI SNFs relative to the comparison SNFs. In addition, controlling for whether episodes were discharged from certified Comprehensive Stroke Centers in the risk-adjustment regression models did not change the impact estimates.

[74] After adding the days between assessments to the models, the DiD estimates became closer to zero, but three out of the five results remained statistically significant at the 10% level.

Exhibit 65: Impact of BPCI on Claim- and Assessment-based Quality Outcomes, by Clinical Episode, Model 3 SNF, Baseline to Intervention, Q4 2013 – Q4 2016

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	All-cause Mortality Rate, 90 days from episode start date (pp)	ED Use, 90 days from episode start date (pp)	Unplanned Readmission Rate, 90 days from episode start date (pp)	Successful Discharge to the Community (pp)	Improved Mobility Function, ADL (pp)	Improved Overall Function, ADL (pp)	Improved Self-care Function, ADL (pp)
Congestive heart failure	2,977	-0.9	1.2	-0.2	-1.5	-2.4	0.5	1.0
COPD, bronchitis, asthma	903	4.0	4.0*	3.7	-4.6	-1.2	0.6	-0.2
Hip and femur procedures except major joint	2,440	0.3	0.6*	2.4	1.8	-6.4	-7.2	-6.8
Major joint replacement of the lower extremity	6,615	-1.0^	0.7	-0.3	-0.8	-5.0	-3.2	-6.7
Medical non-infectious orthopedic	2,247	-2.2	0.1	1.8	-1.8	5.1	2.6	4.3
Other respiratory	727	1.9*	4.2	3.1	-0.8	0.4	-2.6	0.5
Renal failure	1,421	1.8	-0.6	-3.1	-3.5	8.3	4.0	-0.4
Sepsis	4,471	-0.4	2.3	-3.0^	0.8	0.3	-2.1	-0.7
Simple pneumonia and respiratory infections	2,919	1.0	-0.2	-0.4	0.2*	-0.6	-0.7	-0.9
Stroke	1,365	-1.4*	3.2	4.4^	-6.0	-3.7	-0.3	-2.2
Urinary tract infection	2,036	2.3	0.5	2.0*	-2.4	-0.5	-0.6	0.4

* This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate. Equal trends test was conducted for total amount paid by Medicare IP through 90-day PDP, total amount included in the bundle definition for 90-day episodes, emergency department visits, readmission, and mortality outcomes.

^ The sensitivity test suggests that the statistically significant result may be due to the random selection of comparison episodes in the matched sample. See **Appendix D** for additional information on the sensitivity test methodology. Results of the sensitivity test are presented with the BPCI impact estimates in **Appendix H**.

Note: The estimates in this table are the result of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. The successfully discharged to the community measure is populated for approximately 75% of the sample because it does not include patients that were not discharged to the community (i.e., they were discharged to a hospital, SNF, IRF, or LTCH). The assessment outcomes are populated for approximately 80% of the results, ranging from 73% to 88% across outcomes and clinical episodes. SNF = skilled nursing facility. pp = percentage points. ED = emergency department. ADL = activities of daily living. COPD=chronic obstructive pulmonary disease.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

2. Patient Functional Status and Health Care Experience

This section presents patient care experiences and functional outcomes from the beneficiary survey from May 2015 through June 2017.

a. Key Findings

- BPCI respondents with MJRLE and hip and femur procedures except major joint indicated statistically significantly better changes in self-reported functional status from before hospitalization to after the SNF stay, than did comparison respondents. We did not find any difference in functional status between BPCI and comparison respondents with sepsis episodes.
- Relative to the comparison group, a smaller proportion of BPCI respondents with MJRLE reported favorable care experiences than comparison respondents, while BPCI respondents with hip and femur episodes reported better care experience than comparison respondents. There was no difference on measures of care experience between BPCI and comparison respondents with sepsis.

b. Methods

The same methods described in Section II.B.2 above were used in the analyses described in this chapter, with some exceptions noted. The survey sample for BPCI Model 3 SNF episodes included three clinical episodes because other clinical episodes did not have sufficient volume to power an analysis. These three episodes, MJRLE, hip and femur procedures except major joint, and sepsis, were three of the four the largest in volume, together accounting for roughly one third of all Model 3 SNF episodes. We did not pool results across Model 3 SNF clinical episodes as we did for hospital and PGP EIs under Model 2, because the three clinical episodes only represent 34% of total volume for Model 3 SNF EIs, and results from these three strata cannot be interpreted as a summary measure. See **Appendix D** for more detail about the survey measures, sample selection, and other methods.

c. Results

Results for the three Model 3 SNF clinical episodes can be found in **Appendix I**. For survey respondents with MJRLE, BPCI respondents were statistically significantly more likely than comparison respondents to report improvement in use of a mobility device (2.98 percentage points: 47.84 BPCI vs. 44.86 comparison; $p<0.10$) and less likely to report a decline in walking without rest (-3.83 percentage points: 24.66 BPCI vs. 28.49 comparison; $p<0.05$). These results are opposite in direction to the ADL mobility measure reported in Exhibit 65, however, the findings are not inconsistent because the measures and time periods differ. Although remaining differences were not statistically significant, all other rates of improvement were greater and all rates of decline were lower among BPCI respondents than comparisons. This suggests a positive relationship between functional status and BPCI for MJRLE patients treated at SNF EIs. The proportion of respondents with favorable care experiences was slightly smaller for BPCI survey respondents than comparison respondents for two of nine measures: medical staff clearly explained how to take medications before going home (-2.03 percentage points: 90.07 BPCI vs. 92.09 comparison; $p<0.10$) and medical staff clearly explained what follow-up appointments or treatments would be needed before going home (-2.56 percentage points: 91.03 BPCI vs. 93.59 comparison; $p<0.05$).

Among survey respondents with hip and femur episodes, BPCI respondents were less likely to report a decline in walking without rest than were their comparison counterparts (-7.28 percentage points: 66.23 BPCI vs. 73.51 comparison; $p<0.05$). As noted for MJRLE, this finding is not inconsistent with the relative decline in the Activities of Daily Living (ADL) mobility measure for hip and femur episodes reported in Exhibit 65 above due to differences in the measure and time period. Although this was the only statistically significant difference, there was a clear pattern of results favoring BPCI respondents across measures of functional status. BPCI respondents with hip and femur episodes reported better experiences on all nine care experience measures than did comparisons respondents, although only one was statistically significant (less likely to receive conflicting medical advice: +6.10 percentage points: 67.04 BPCI vs. 60.94 comparison; $p<0.10$). However there was a statistically significant difference in response rates (46.55 for BPCI vs. 41.89 for comparison, $p<0.05$) with the comparison group less likely to respond, so the results for this clinical episode should be interpreted with caution.

There was no consistent pattern of changes in functional status, among BPCI respondents with sepsis, or any pattern of better or worse care experience. However, as with hip and femur, results may be biased by the statistically significant difference in response rates for the BPCI and comparison groups.

3. Understanding the Impact of Terminating Participation on Payments

The BPCI initiative allows participants to withdraw from a clinical episode at any time.[75] By December 2016, 21% of the Model 3 SNF EIs withdrew from at least one clinical episode.[76] Descriptive analysis suggests that SNF EIs were more likely to withdraw from clinical episodes in which they had negative NPRA, and this non-random withdrawal could affect our BPCI impact estimates by disrupting the balance between the BPCI and comparison groups, as well as giving more weight to participants with a longer exposure to BPCI. It is not clear how these potential biases would affect the BPCI impact estimates. To evaluate the impact of withdrawals on the estimates, we re-calculated the estimates with all episodes from participants and comparison group providers included in the analysis through December 2016, regardless of whether the BPCI EI had withdrawn from the clinical episode prior to December 2016. We refer to these sensitivity analyses as ITT DiD analyses.

Across all clinical episodes, 15% of the matched SNF EI episodes were initiated after the EI had withdrawn from the clinical episode. Across clinical episodes, the share of matched post-withdrawal episodes as a proportion of the intervention sample varied widely from 2% for hip and femur procedures except major joint to 31% for stroke. Results by clinical episode are located in **Appendix K**.

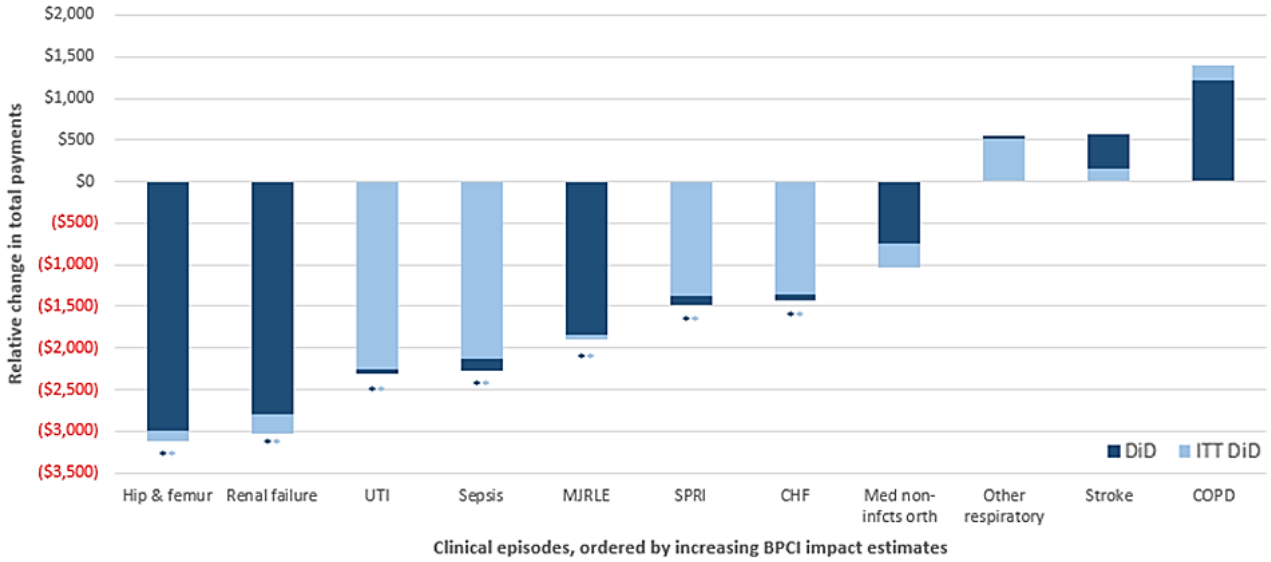
All seven Model 3 SNF clinical episodes that exhibited a statistically significant relative decline in total allowed payments included in the bundle definition had statistically significant ITT DiD estimates in the same direction (Exhibit 66). This indicates that BPCI participants achieved a

[75] We use the term withdrawal to reflect EIs that terminate their participation in BPCI or stop participating in the given clinical episode. Note that EIs that participated in a clinical episode for only one quarter are excluded from both the BPCI impact analysis and this analysis.

[76] This is limited to participants that met the minimum number of patient episodes to be included in our DiD and ITT DiD analysis.

decline in total allowed payment amounts relative to the comparison group for these Model 3 SNF clinical episodes, suggesting that the BPCI impact estimates are robust to selective withdrawal. For additional results and methods, see **Appendix K**.

Exhibit 66: BPCI Impact Estimates and ITT DiD Estimates for Total Medicare Allowed Amounts Included in the Bundle Definition, by Clinical Episode, Model 3 SNF, Baseline to Intervention, Q4 2013 – Q4 2016



Note: The estimates in this exhibit are the results of the BPCI impact difference-in-differences (DiD) model (dark blue) and intent-to-treat (ITT) DiD model (light blue). The results are ranked by increasing magnitude of the BPCI impact estimates. BPCI impact estimates that are significant at the 10% significance level are indicated by a dark blue asterisk; ITT DiD estimates that are significant at the 10% significance level are indicated by a light blue asterisk. The Medicare payment outcome is standardized to remove the effect of geographic and other adjustments. The total amount included in the bundle definition is based on only the 90-day episodes. SNF=skilled nursing facility. UTI = urinary tract infection. MJRLE = major joint replacement of the lower extremity. SPRI = simple pneumonia and respiratory infections. CHF = congestive heart failure. Med non-infcts orth = medical non-infetious orthopedic. COPD = chronic obstructive pulmonary disease.

***Source*:** Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

4. Market Dynamics

This section presents the analysis of market share and provider referral patterns for Model 3 SNF participants.

a. Key findings

- Overall, there was no change in the market share from the baseline period to the BPCI period. Results varied considerably across clinical episodes and EI cohorts.

b. Methods

We used the same analytical approach that was used to analyze Model 2 hospital EIs, described in Section II.B.5 above. Details about the analytical methods can be found in **Appendix D**.

c. Results

Overall, the median market share was 2.3% in the baseline period. There was a negligible median change of 0.01 pp in the market share from the baseline period to the last BPCI year in this analysis (Exhibit 67). However, results varied by clinical episode and cohort, as well as by cohort within the same clinical episode, e.g., SPRI (see **Appendix L** for results by clinical episode and cohort).

Exhibit 67: Median Baseline Rate and Median Changes from Baseline for Market Share, Model 3 SNF Episode Initiators, Q4 2011-Q4 2016

Measures	Median Baseline Rate (%)	Median Changes from Baseline (pp)
EI Market Share	2.3	0.01

Note: This table shows the measure rate in the baseline period, and the changes from the baseline period to the last BPCI year in the analysis. SNF = skilled nursing facility. pp=percentage points.

Source: Lewin analysis of Medicare claims for discharges that began Q4 2011 through Q4 2016.

C. Impact of BPCI among Participating HHAs

1. Payment, Utilization, Quality, and Patient Mix

This section presents the BPCI impact estimates on payments, utilization, quality, and characteristics of the patient mix for episodes initiated by home health agencies for the first 13 quarters of the initiative, Q4 2013 through Q4 2016.

a. Key Findings

- Among the three HHA-initiated clinical episodes with sufficient sample size, there was one clinical episode with a statistically significant reduction in the total payment amount included in the bundle definition for 90-day episodes.
- There were few statistically significant relative changes in utilization or quality across any of the three clinical episodes.

b. Methods

The same methods described in Section II.B.1 above were used in the analyses described in this section. We constructed comparison groups for three Model 3 HHA clinical episodes that were deemed to have a sufficient sample size for meaningful analysis. The episodes in these three clinical episodes represent 61% of all episodes initiated by Model 3 HHA EIs during the first 13 quarters of the initiative. See **Appendix D** for a more detailed description of the methodology.

c. Results

Sample characteristics

The number of Model 3 HHA EIs included in the sample ranged from 37 to 46 across the three clinical episodes for which there was sufficient sample size to analyze (Exhibit 68). Because providers could join BPCI over an extended period and could terminate participation in a clinical episode at any given time or withdraw from the initiative completely, these data represent an

average of six quarters of participation from Q4 2013 to Q4 2016. During the first thirteen quarters, 35 of the 116 HHA EIs (30%) withdrew entirely from the initiative. The percent of EIs that exited varies across the three HHA clinical episodes. Among the 71 HHAs participating in any of the three clinical episodes analyzed, approximately 46% stopped participating in at least one clinical episode by the end of Q4 2016. The BPCI impact estimates include episodes from all HHA EIs during their participation period. The contribution of episodes from HHA EIs that stopped participating by Q4 2016 varies by clinical episode. Over half of the HHA EIs stopped participating in the CHF clinical episode, and these EIs accounted for 19% of intervention episodes. For simple pneumonia and respiratory infections, terminated EIs (57%) accounted for 55% of the episodes initiated during the intervention period. Retention was higher in the MJRLE clinical episode, where only five (14%) EIs stopped participating by Q4 2016.

Exhibit 68: Characteristics of the Matched BPCI Providers Included in the BPCI Impact Estimates, Model 3 HHA, Q4 2013 – Q4 2016

Clinical Episode	BPCI HHAs (#)	Matched Intervention Period Episodes (#)	Average Length of Participation (Quarters)	EIs that Terminated Participation in the Clinical Episode (#)	Episodes from EIs that Terminated (%)
Congestive heart failure	46	4,492	5	26	19.3
Major joint replacement of the lower extremity	37	3,541	6	5	4.5
Simple pneumonia and respiratory infections	37	1,273	6	21	55.1

Note: The sample does not include BPCI providers for which no comparison provider was found; approximately 77% of BPCI HHAs were matched to comparison HHAs. See **Appendix D** for information on the methods used to determine the sample. DiD=difference-in-differences. HHA=home health agency. EI=episode initiator. This exhibit is limited to the matched BPCI providers used to calculate the DiD results in the reminder of this section.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2013 through Q4 2016 for BPCI providers.

Has patient mix changed under BPCI?

Because target prices are not risk adjusted, episode payments for BPCI participants could decline relative to their target price if their mix of patients during the intervention period is less resource intensive than their mix of patients during the baseline. Based on claim- and assessment-based characteristics that are associated with higher resource use, there is evidence that the resource intensity of CHF episodes declined for BPCI-participating HHAs from the baseline to the intervention period relative to the comparison group. Five of patient assessment measures from the Outcome and Assessment Information Set (OASIS) indicated a less resource intensive patient mix and were statistically significant ($p<0.05$). Among SPRI patients, four of the OASIS measures as well as two of the claim-based measures indicated a more intensive patient mix and were statistically significant ($p<0.10$). There was no consistent pattern among MJRLE patients. See Exhibits 69 and 70 for results of the changes in patient mix in claim- and assessment-based characteristics. See **Appendix D** for additional information on measure definitions, categorization of clinical episodes, and other methods.

Exhibit 69: Relative Changes in Claim-based Characteristics, by Clinical Episode Strata, Baseline to Intervention, Model 3 HHA, Q4 2013 – Q4 2016

Change in Patient Mix	Clinical Episode Strata	Number of Episodes Q4 2013- Q4 2016	Age: 80+ years	Medicaid Eligibility	Disabled, no ESRD	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
Less resource intensive	Congestive heart failure	4,492	-2.2	4.4	0.8	0.04	0.5	-1.4	0.3	-0.3
More resource intensive	Simple pneumonia and respiratory infections	1,273	9.6	-1.7	0.1	0.08	1.1	2.0	6.8	-3.2
No consistent pattern	Major joint replacement of the lower extremity	3,541	1.4	-0.8	-0.5	0.02	2.3	1.9	0.8	-2.9

* These characteristics measure utilization of care in the six months prior to the anchor hospitalization. Count of HCCs is based on the six months prior to the anchor hospitalization.

Note: Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Categorization of resource intensity was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates. See Appendix D for additional details of the categorization. HHA = home health agency. ESRD=end-stage renal disease. HCC=hierarchical conditions categories.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q4 2016 (intervention period) for BPCI EIs and the matched comparison providers.

Exhibit 70: Relative Changes of Assessment-based Characteristics of Beneficiaries, by Clinical Episode Strata, Baseline to Intervention, Model 3 HHA, Q4 2013 – Q4 2016

Change in Patient Mix	Clinical Episode Strata	Number of Episodes Q4 2013 – Q4 2016	Poor Overall Status	Impaired Vision or Hearing	Impaired Cognition	Unhealed Pressure Ulcer	Short of Breath from Moderate to No Exertion	Require Use of Bedside Commode or are Totally Dependent in Toileting	Require Assistance Transferring or are Unable to Transfer (e.g. from bed to wheelchair)	Require Walker or More Assistance Ambulating	Dependent in Maintaining Self-care	Dependent in Ambulating	Dependent in Transferring	Not Likely to Receive Assistance in ADL	Caregiver Needs Training to Provide Supervision & Safety, is Unlikely to Provide Help, or is not Present	Incontinence	Depressive Symptoms
Less resource intensive	Congestive heart failure	4,492	-10.7	-0.3	0.6	0.5	-4.1	-6.9	-29.8	-3.7	-1.2	0.3	-0.4	1.7	0.4	2.1	0.3
More resource intensive	Simple pneumonia and respiratory infections	1,273	0.2	2.2	1.8	0.3	1.2	4.4	3.2	1.6	4.2	3.2	3.2	0.5	-0.3	0.9	-1.8
No consistent pattern	Major joint replacement of the lower extremity	3,541	0.3	0.5	1.6	0.4	2.9	-8.6	-5.7	0.8	0.1	0.8	0.3	1.5	1.1	-3.3	0.7

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. The Outcome and Assessment Information Set (OASIS) assessment is administered within five days of the start of care date. The measurement period for outcomes listed above is upon assessment, except for depressive symptoms, for which the measurement period covers the 14 days prior to the assessment date. Assessment data was not available for all episodes. This table is limited to the episodes where we had the initial patient assessment data. Categorization of resource intensity was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates. See Appendix D for additional details of the categorization. HHA=home health agency. ADL=Activities of Daily Living.

Source: Lewin analysis of Medicare claims, enrollment, and OASIS data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q4 2016 (intervention period) for BPCI EIs and the matched comparison providers.

How have average standardized payments changed under BPCI?

The total allowed payment amount (Medicare program payments plus coinsurance and/or copayments) included in the bundle declined from baseline to intervention for BPCI HHAs episodes relative to the comparison group in two of the three clinical episodes, and the decline was statistically significant in one ($p<0.10$) (Exhibit 71). There were no statistically significant relative changes in SNF or HHA payments during the 90-day PDP in any of the three clinical episodes. There was a statistically significant relative increase in the standardized allowed amount for readmissions for one clinical episode ($p<0.10$). Detailed results of the BPCI impact estimates by clinical episode are located in **Appendix P**.

Exhibit 71: Impact of BPCI on Payment Outcomes, by Clinical Episode, Model 3 HHA, Baseline to Intervention, Q4 2013-Q4 2016

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	Total Amount Included in Bundle Definition[1]	Total Amount Paid by Medicare, IP through 90-day PDP[2]	HHA Standardized Allowed Amount, 90-day PDP[3]	SNF Standardized Allowed Amount, 90-day PDP[3]	Readmissions Standardized Allowed Amount, 90-day PDP[3]	Part B Standardized Allowed Amount, 90-day PDP	Total Part A & B 30-day Post-bundle Period
Congestive heart failure	4,492	-$791	-$785	-$75	-$153	-$291	-$158	**-$591**
Major joint replacement of the lower extremity	3,541	-$512*	$415	-$116	-$17	-$99	$2	$258
Simple pneumonia and respiratory infections	1,273	$1,085	**$2,031**	$37	$103	$812	$39	$71

[1]The total amount included in bundle definition values are based on only the 90-day episodes.

[2]Total amount paid does not include beneficiary out-of-pocket expenses.

[3]These payment measures are not conditional upon use of the service.

* This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate. Equal trends test was conducted for total amount paid by Medicare IP through 90-day PDP, total amount included in the bundle definition for 90-day episodes, emergency department visits, readmission, and mortality outcomes.

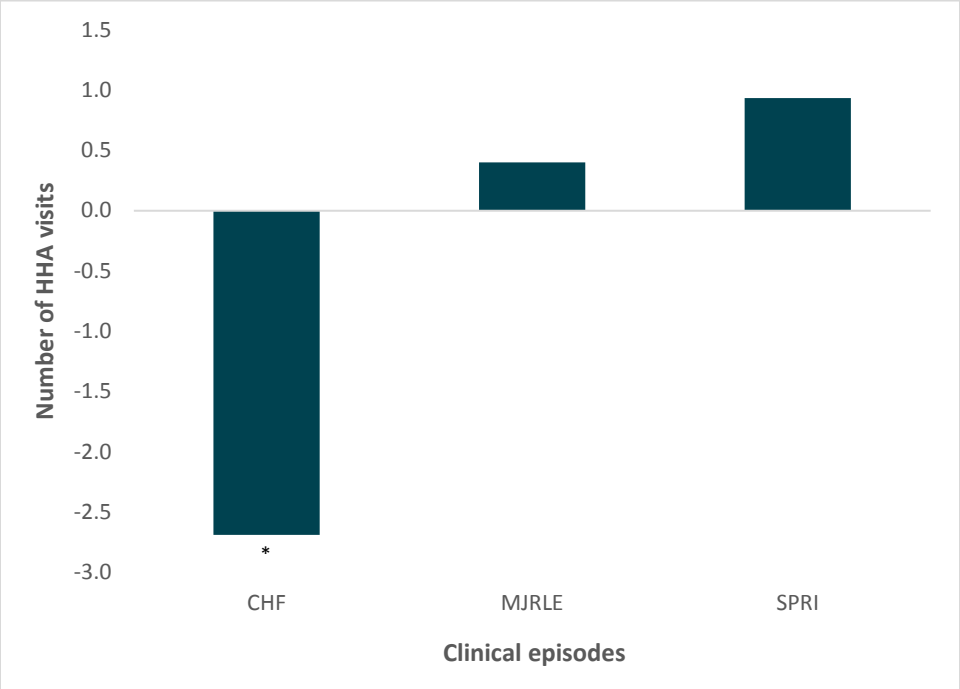
Note: The estimates in this table are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented, either due to insufficient sample size or the type of episodes initiated during the time period. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments and are trended to 2015. HHA = home health agency. PDP = post-discharge period. IP = in-patient. SNF = skilled nursing facility.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

How have the services changed under BPCI?

There was a statistically significant decline in the number of HHA visits relative to the change in the comparison group for one clinical episode (Exhibit 72).

Exhibit 72: Impact of BPCI on Home Health Visits, by Clinical Episode, Model 3 HHA, Baseline to Intervention, Q4 2013 – Q4 2016



*Indicates DiD estimates are statistically significant at the 10% level.

Note: The estimates in this table are the results of a difference-in-differences (DiD) model. HHA = home health agency. CHF = congestive heart failure. MJRLE = major joint replacement of the lower extremity. SPRI = simple pneumonia and respiratory infection.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

How has quality of care changed under BPCI?

Claim-based results suggest that the quality of care did not change under BPCI Model 3 HHAs (Exhibit 73). Across clinical episodes, there were no patterns in the change in readmissions rates, emergency department (ED) use, or mortality rates in the 90 days from the episode start date. For CHF episodes, the mortality rate declined from baseline to intervention relative to the change in the comparison group ($p<0.10$). However, an additional sensitivity test revealed that this decline may be due to a relatively less resource intensive patient mix as observed in the initial patient assessments.[77] There was a statistically significant decline relative to the comparison group in ED use for simple pneumonia and respiratory infections episodes ($p<0.10$). There was a statistically significant decline in readmission rates for MJRLE episodes ($p<0.10$), but the sensitivity results suggest this may have been a result of the random episode selection from comparison HHAs, and it may not have been statistically significant if a different sample of

[77] We added the variables from the OASIS initial assessments in the patient mix tables in Exhibit 73 above to the risk-adjustment regression models and compared the DiD estimates to the original BPCI impact estimates.

comparison episodes had been selected. (Results of this sensitivity test are presented with the BPCI impact estimates in **Appendix H**.)

The assessment-based quality outcomes also suggest that the quality of care did not change among the approximately three-quarters of patients who have two patient assessments (Exhibit 73). Only one assessment-based outcome was statistically significant: there was a decrease in the proportion of patients with improved bed transferring relative to the change in the comparison group for simple pneumonia and respiratory infections episodes ($p<0.10$).

Exhibit 73: Impact of BPCI on Claim- and Assessment-based Quality Outcomes, by Clinical Episode, Model 3 HHA, Baseline to Intervention, Q4 2013 – Q4 2016

Clinical Episode	Number of Episodes Q4 2013 - Q4 2016	All-cause Mortality Rate, 90 days from the episode start date (pp)	ED Use, 90 days from the episode start date (pp)	Unplanned Readmission Rate, 90 days from the episode start date (pp)	Improved Ambulation, ADL (pp)	Improved Bathing, ADL (pp)	Improved Bed Transferring, ADL (pp)	Improved Lower Body Dressing, ADL (pp)	Improved Upper Body Dressing, ADL (pp)
Congestive heart failure	4,492	-1.8	1.7	-0.9	-2.7	2.9	0.1	1.8	3.3
Major joint replacement of the lower extremity	3,541	0.1	0.0	-1.6*^	-1.0	0.0	1.4	-1.9	-0.9
Simple pneumonia and respiratory infections	1,273	-0.3	-4.8	3.5*	-3.6	-0.1	-5.7	-4.7	-4.2

* This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate. Equal trends test was conducted for total amount paid by Medicare IP through 90-day PDP, total amount included in the bundle definition for 90-day episodes, emergency department visits, readmission, and mortality outcomes.

^ The sensitivity test suggests that the statistically significant result may be due to the random selection of comparison episodes in the matched sample. See **Appendix D** for additional information on the sensitivity test methodology. Results of the sensitivity test are presented with the BPCI impact estimates in **Appendix H**.

Note: The estimates in this table are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented, either due to insufficient sample size. The assessment outcomes are populated for approximately 75% of the results, ranging from 71% to 85% across outcomes and clinical episodes. HHA=home health agency. pp=percentage points. ED=emergency department. ADL=Activities of Daily Living.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

2. Patient Functional Status and Health Care Experience

This section presents patient care experiences and functional outcomes from the beneficiary survey from May 2017 through June 2017.

a. Key Findings

- BPCI Model 3 HHA respondents with MJRLE and CHF episodes reported statistically significantly worse average changes in functional status relative to comparison respondents.
- BPCI respondents with MJRLE also reported statistically significantly worse care experience relative to comparison respondents.

b. Methods

The same methods described in Section II.B.2 were used in the analyses described in this chapter, with some exceptions noted. The survey sample for BPCI Model 3 HHA included only two clinical episodes, because other clinical episodes did not have sufficient volume to power an analysis. These two episodes are the largest in volume and together account for approximately half of all BPCI episodes initiated in HHAs. See **Appendix D** for more detail about the survey measures, sample selection, and other methods.

c. Results

Results for the two Model 3 HHA clinical episodes can be found in **Appendix I**. BPCI respondents with CHF were statistically significantly more likely than comparison respondents to report a decline in walking without rest (7.24 percentage points: 47.73 BPCI vs. 40.50 comparison; $p<0.05$). Most of the other functional status measures were negative but not statistically significant. BPCI respondents in this clinical episode were statistically significantly less likely than comparison respondents to report that medical staff always spoke to them in their preferred language (-8.17 percentage points: 82.70 BPCI vs. 90.87 comparison; $p<0.05$).[78] They were statistically significantly more likely to say that they were discharged from the hospital at the right time (5.46 percentage points: 89.41 BPCI vs. 83.95 comparison; $p<0.05$).

BPCI respondents with MJRLE reported statistically significantly lower rates of improvement in using a mobility device (-3.94 percentage points: 64.20 BPCI vs. 68.13 comparison; $p<0.10$) and in planning regular tasks (-3.95 percentage points: 80.44 BPCI vs. 84.39 comparison; $p<0.05$) than did comparison respondents, although planning tasks may not be especially salient for patients with MJRLE episodes. BPCI respondents also reported statistically significantly higher rates of decline in using a mobility device (3.13 percentage points: 23.92 BPCI vs. 20.78 comparison $p<0.10$). BPCI respondents with MJRLE also reported statistically significantly worse care experience for two measures: they were more likely to report receiving conflicting medical advice from staff (-4.32 percentage points: 82.44 BPCI vs. 86.76 comparison; $p<0.05$), and less likely to report that services were appropriate for their level of care (-6.24 percentage points: 70.41 BPCI vs. 76.65 comparison; $p<0.05$). Despite these care experiences, BPCI respondents reported similar overall satisfaction with their recovery, relative to the comparison group.

[78] This measure included all survey respondents, but we controlled for English as a preferred language.

3. Understanding the Impact of Terminating Participation on Payments

The BPCI initiative allows participants to withdraw from a clinical episode at any time.[79] By December 2016, 47% of the M3 HHA EIs withdrew from at least one clinical episode.[80] Descriptive analysis suggests that HHA EIs were more likely to withdraw from clinical episodes in which they had negative NPRA, and this non-random withdrawal could affect our BPCI impact estimates by disrupting the balance between the BPCI and comparison groups, as well as giving more weight to participants with a longer exposure to BPCI. It is not clear how these potential biases would affect the BPCI impact estimates. To evaluate the impact of the withdrawals on the BPCI impact estimates, we re-calculated the estimates with all episodes from the participants and comparison group providers included in the analysis through December 2016, regardless of whether the BPCI EI had withdrawn from the clinical episode prior to December 2016. We refer to these sensitivity analyses as ITT DiD analyses.

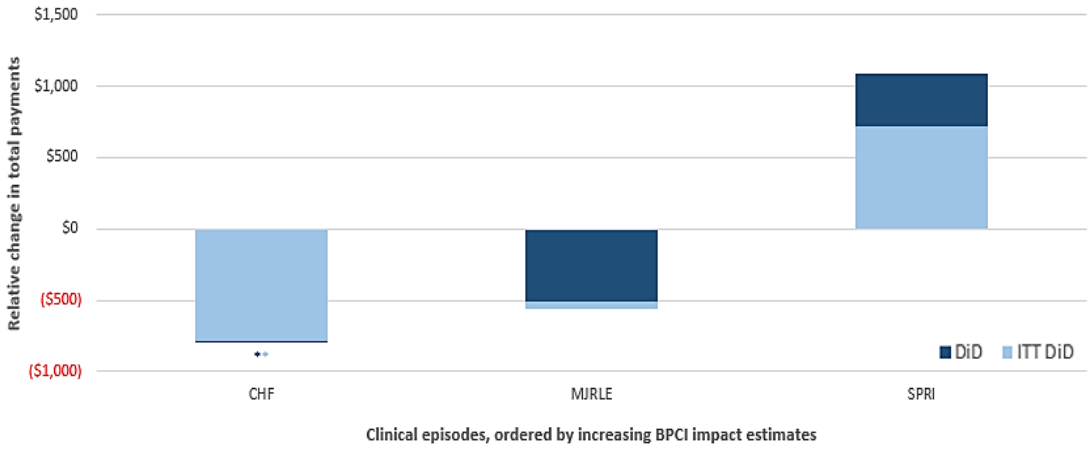
Across all clinical episodes, 23% of the matched HHA EI episodes were initiated after the EI had withdrawn from the clinical episode. Across clinical episodes, the share of matched post-withdrawal episodes as a proportion of the intervention sample varied widely: 1% of major joint replacement of the lower extremity episodes were initiated post-withdrawal compared to 25% of CHF and 40% of simple pneumonia and respiratory infections.

The Model 3 HHA clinical episode that exhibited a statistically significant relative decline in total allowed payments included in the bundle definition, CHF, also had a statistically significant ITT DiD estimate in the same direction (Exhibit 74). This indicates that despite selective withdrawal, BPCI participants in CHF achieved a decline in total allowed payment amounts relative to the comparison group, suggesting that the BPCI impact estimate is robust to selective withdrawal. For additional results and methods, see **Appendix K**.

[79] We use the term withdrawal to reflect EIs that terminate their participation in BPCI or stop participating in the given clinical episode. Note that EIs that participated in a clinical episode for only one quarter are excluded from both the BPCI impact analysis and this analysis.

[80] This is limited to participants that met the minimum number of patient episodes to be included in our DiD and ITT DiD analysis.

Exhibit 74: BPCI Impact Estimates and ITT DiD Estimates for Total Medicare Allowed Amounts Included in the Bundle Definition, by Clinical Episode, Model 3 HHA, Baseline to Intervention, Q4 2013 – Q4 2016



Note: The estimates in this exhibit are the results of the BPCI impact difference-in-differences (DiD) model (dark blue) and intent-to-treat (ITT) DiD model (light blue). The results are ranked by increasing magnitude of the BPCI impact estimates. BPCI impact estimates that are significant at the 10% significance level are indicated by a dark blue asterisk; ITT DiD estimates that are significant at the 10% significance level are indicated by a light blue asterisk. The Medicare payment outcome is standardized to remove the effect of geographic and other adjustments. The total amount included in the bundle definition is based on only the 90-day episodes. HHA=home health agency. CHF=congestive heart failure. MJRLE=major joint replacement of the lower extremity. SPRI=simple pneumonia and respiratory infections.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers.

D. Medicare Program Savings

As demonstrated in this evaluation, BPCI participants have successfully reduced Medicare FFS payments across a variety of clinical episodes. Under Model 3, seven of the 11 SNF clinical episodes and one of the three HHA clinical episodes we examined had statistically significant declines in Medicare allowed payments among BPCI participants relative to the comparison group ($p<0.10$) and several others had declines in payments that did not reach statistical significance (see Exhibit 62 in Section III.B.1 and Exhibit 71 in Section III.C.1). However, payment reductions may not translate into net savings to Medicare because they do not account for reconciliation payments, including net payment reconciliation amounts (NPRA) and other amounts that Medicare paid (e.g., outlier payments) or recovered from participants. In this section, we first estimate the total change in non-standardized payments,[81] that is, the actual provider payments from Medicare, for all of the Model 3 clinical episodes with sufficient volume. Then, we compute net savings to Medicare from Q4 2013 to Q4 2016 by subtracting reconciliation payments from the total change in non-

[81] Non-standardized payments vary from the standardized allowed amounts that we use in the BPCI impact analyses. Non-standardized amounts include adjustments for geographic differences in wages and other costs, teaching, and treating a disproportionate share of low-income beneficiaries. These adjustments are removed in the standardized amounts so that BPCI impact analyses can isolate the effect of BPCI.

standardized payments attributable to BPCI. We also calculate the net savings to Medicare for a scenario in which CMS had not eliminated negative NPRA repayment retrospectively.[82]

1. Key Findings

- The DiD estimate for the total change in non-standardized payments indicated a statistically significant ($p<0.10$) reduction in payments due to BPCI for Model 3 episodes ($85.4 million overall, $924 per episode).
- After accounting for the $170.5 million in reconciliation payments for Model 3 EIs through Q4 2016, BPCI resulted in a net loss to Medicare of $85.2 million (-$921 per episode) ($p<0.10$).
- Had CMS not eliminated downside risk, reconciliation payments would have been $140.6 million, and the net loss to Medicare would have been $55.2 million (-$597 per episode) for BPCI Model 3 through Q4 2016 ($p<0.10$).

2. Methods

The net savings to Medicare for BPCI Model 3 from Q4 2013 through Q4 2016 was calculated using the following formula:

Net savings to Medicare = DiD estimate of total change in non-standardized payments – reconciliation payments

See **Appendix D** for additional details on the definitions and calculations for each component.

3. Results

The results of the DiD analysis showed that episodes initiated by Model 3 SNFs and HHAs were associated with a statistically significant average decline in payments of $872 per episode ($p<0.10$). After multiplying the DiD estimate of change in per-episode standardized payments by the total number of Model 3 BPCI episodes and adjusting for standardization, the estimate of total change in non-standardized payments was $85,358,425 across all Model 3 episodes.

Reconciliation payments through Q4 2016 were $170,524,351. A majority of this amount ($159.2 million) was due to NPRA payments. After subtracting reconciliation payments from the DiD estimate of total change in non-standardized payments, we estimate that BPCI resulted in a net loss to Medicare of $85,165,926 over the first 13 quarters of the initiative (Exhibit 75). The loss was statistically significant ($p<0.10$).

[82] In November 2014, CMS eliminated downside risk for episodes initiated between Q4 2013 and Q4 2014 due to inaccurate target prices. In July 2016, CMS eliminated downside risk for any episode of care that was initiated as a result of the episode attribution issues caused by incorrect PGP Reassignment Lists initiated in 2015, and in December 2017, CMS extended this elimination of downside risk for all such episodes initiated through Q3 2016.

Exhibit 75: Components of Net Savings to Medicare, Model 3, Q4 2013 – Q4 2016

Component	Estimate	LCI	UCI
Total change in non-standardized payments	$85,358,425	$44,931,880	$125,784,971
Reconciliation payments	-$170,524,351	–	–
Net savings to Medicare	-$85,165,926	-$125,592,471	-$44,739,380

Note: The estimates of total change in non-standardized payments in this table are from a difference-in-differences (DiD) model of standardized Medicare paid amounts during the qualifying inpatient stay and 90 day post-discharge period. The coefficients from the DiD estimate of change in per-episode standardized payments is multiplied by -1, multiplied by the number of Model 3 intervention episodes, and converted to non-standardized amounts. Reconciliation payments include NPRA payments and other payments to BPCI-participating providers. Net savings to Medicare is the difference between the total change in non-standardized payments and reconciliation payments. LCI = lower 90% confidence interval. UCI = upper 90% confidence interval.

***Source*:** Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers. CMS data on reconciliation payments.

If CMS had not eliminated downside risk, Medicare would have recovered an additional $29,949,796 of NPRA. The difference between actual and planned NPRA was $7,753,007 for SNF participants, $3,361,997 for HHA participants, $16,525,727 for PGP participants, $1,426,997 for IRF participants, and $882,069 for LTCH participants.[83] Subtracting the reconciliation payments of $140,574,555 from the total change in non-standardized payments gives an estimated net loss to Medicare of $55,216,130 (Exhibit 76). This suggests there would have been a net loss to Medicare even if CMS had imposed downside risk for some episodes, though the loss would have been cut by approximately $30 million.

Exhibit 76: Components of Net Savings to Medicare if Downside Risk Had Not Been Eliminated, Model 3, Q4 2013 – Q4 2016

Component	Estimate	LCI	UCI
Total change in non-standardized payments	$85,358,425	$44,931,880	$125,784,971
Reconciliation payments (downside risk not eliminated)	-$140,574,555	–	–
Net savings to Medicare (downside risk not eliminated)	-$55,216,130	-$95,642,675	-$14,789,584

Note: The estimates of total change in non-standardized payments in this table are from a difference-in-differences (DiD) model of standardized Medicare paid amounts during the qualifying inpatient stay and 90 day post-discharge period. The coefficients from the DiD estimate of change in per-episode standardized payments is multiplied by -1, multiplied by the number of Model 3 intervention episodes, and converted to non-standardized amounts. Reconciliation payments (downside risk not eliminated) include net NPRA payments and other payments to BPCI-participating providers, adjusted for the difference between actual and planned (downside risk not eliminated) NPRA. Net savings to Medicare (downside risk not eliminated) is the difference between the total change in non-standardized payments and reconciliation payments (downside risk not eliminated). LCI = lower 90% confidence interval. UCI = upper 90% confidence interval.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q4 2016 for BPCI and comparison providers. CMS data on reconciliation payments.

E. Model 3 Discussion

There were 873 SNF, 144 PGP, 116 HHA, 9 IRF, and 1 LTCH episode initiators (EIs) active in Model 3 during the first 13 quarters of BPCI. We had sufficient sample size to conduct impact analyses on 11 SNF and 3 HHA clinical episodes. Our study sample included 493 SNFs and 71 HHA EIs that initiated 28,121 and 9,306 episodes of care, respectively, during their tenure in the BPCI initiative. Although Model 3 participants achieved statistically significant relative reductions

[83] The impact of eliminating downside risk for the first five quarters of the initiative was $13,424,070. The impact of continuing to eliminate downside risk for PGPs until Q3 2016 was $16,525,727.

in total Medicare allowed amounts for 8 out of the 14 clinical episodes we could examine, our analyses indicated that Medicare costs increased as a result of Model 3 when NPRA and other costs were considered. Further, even though claim-based outcomes indicated no declines in the quality of care, patient-reported relative changes in functional status and care experiences were mixed across SNF and HHA EIs and clinical episodes.

The SNFs and HHAs that elected to participate in BPCI differed from those that did not participate in ways that may have contributed to their ability or willingness to engage in the initiative. Participating SNFs and HHAs tended to be larger than the average non-participant, for-profit, and part of a chain. Participants also had higher Medicare Part A payments for their chosen clinical episodes in the pre-BPCI period than providers that did not participate, with only one exception. In addition, virtually all EIs joined BPCI under a convener, or participated as a convener, which could contribute to their ability to succeed under the intervention. Some conveners assumed the financial risk on behalf of their Awardees or EIs. (See **Appendix A** for the definition of each type of convener.) According to participants we interviewed, conveners also offered services that enhanced care redesign, informed data-driven decisions, and eased administrative burden.

Few Model 3 participants used waivers from certain Medicare program requirements and fraud and abuse laws during the first 13 quarters of the initiative. Some Awardees obtained the incentive waiver to provide equipment to help beneficiaries monitor their medical condition, but fewer than 600 beneficiaries received an incentive during this period. Relatively few Model 3 Awardees entered into gainsharing agreements. By March 2017, 17 Awardees distributed $19.8 million to their gainsharing partners, which constitutes 21% of the NPRA earned by all Model 3 Awardees through Q3 2016. These agreements were typically among PAC providers in the same BPCI convener group; Awardees said there were significant barriers to engaging non-affiliated hospitals and downstream PAC providers in gainsharing.

Model 3 SNF EIs achieved reductions in Medicare allowed amounts for some clinical episodes. Sensitivity analyses indicate that the payment reductions were robust and not an artifact of providers leaving the initiative. Similar to what we found for Model 2, episode payments were lowered by reducing institutional PAC. However, because an episode is triggered under Model 3 only with the use of PAC, these participants did not have the option, as participants in Model 2, to forego institutional PAC altogether. Rather, they could only lower PAC payments by reducing SNF length of stay. Across the seven clinical episodes with statistically significant declines in total allowed amounts included in 90-day bundles, the lower payments were due to fewer days in the SNF. At the same time, payments for home health increased in four of these seven SNF clinical episodes. Reducing SNF days, however, may be a risky strategy for BPCI-participating SNFs. For some clinical episodes, the lower SNF payments were partly offset by higher Part B and readmission payments. In addition, fewer SNF days reduced Medicare revenue for these participants. Therefore, a SNF must consider lower Medicare revenue in per diem SNF payments and the risk of higher payments for other services against the promise of positive NPRA under BPCI.

For Model 3 HHA EIs, there were fewer strategies for reducing Medicare allowed amounts. Among the three HHA clinical episodes analyzed, only CHF achieved a statistically significant decline in total payments. Statistically significant declines in post-bundle payments for CHF may indicate that BPCI HHA EIs treated healthier, less costly patients. Though we control for

differences in patient mix that can be observed in the claims data, initial patient assessments for BPCI CHF patients indicate greater improvements in 5 of 15 functional status indicators from baseline to intervention, relative to the change for the comparison group. There was also a statistically significant decline in HHA visits for CHF episodes. Unlike the SNF EIs, which could reduce Medicare payments by reducing the services they provided, reducing HHA visits does not always lower Medicare payments because HHAs typically are paid for a 60-day episode of care.

Indicators of quality of care across Model 3 clinical episodes generally revealed no consistent BPCI impacts. With few exceptions, there were no statistically significant changes in mortality, emergency department use, or readmissions for BPCI relative to the comparison group. Sensitivity tests suggested that the few statistically significant changes, both favorable and unfavorable, were due to the random selection of comparison episodes or to differences in patient mix. Assessment-based quality outcomes indicated that rates of improvement in functional status were lower for BPCI SNF orthopedic surgery episodes, which were partially due to declines in SNF length of stay among the BPCI sample.

Beneficiary survey data indicate some differences in self-reported indicators of quality between BPCI and comparison respondents. BPCI respondents in SNF-initiated MJRLE and hip and femur episodes reported greater rates of improvement and lower rates of decline in mobility than comparison respondents, though few differences were statistically significant. While the direction of results for these two clinical episodes is opposite to what we found for the ADL mobility measure, results are not inconsistent because of differences in the measures and time periods covered. Satisfaction with care experience was mixed among the two orthopedic surgery episodes analyzed. A smaller proportion of BPCI respondents in SNF MJRLE episodes reported favorable care experiences, while a larger proportion of those in hip and femur episodes reported favorable care experiences. Among the two HHA clinical episodes analyzed, BPCI respondents on average reported statistically significantly worse changes in functional status relative to comparison respondents. HHA respondents with a MJRLE episode also reported significantly worse care experience relative to comparison respondents.

There was evidence that patients treated by Model 3 BPCI participants had better functional status upon admission during the intervention period than the baseline for one HHA and five SNF clinical episode strata, relative to the comparison group. It is important to note that SNFs and HHAs have considerable control over the patients they admit, and they may evaluate patients prior to their admission decision. By treating a less costly mix of patients in the intervention period than the baseline, Model 3 EIs may have found it easier to reduce per-episode payments below the target amount and achieve NPRA. These trends of favorable patient selection were observed in both surgical and non-surgical clinical episodes. Favorable patient selection was not evident across all clinical episodes, however. We observed an increase in patient severity for SNF other respiratory episodes and HHA SPRI episodes. In over half the Model 3 clinical episodes, we observed no consistent pattern of changes in patient characteristics.

Overall, there were indications that Model 3 resulted in reductions in episode payments with no consistent or strong indications of changes in quality of care. Further, the statistically significant declines in Medicare allowed payments did not result in Medicare program savings. After accounting for amounts that Medicare paid to or received from participants through NPRA, we estimate a net loss to Medicare of $85.2 million (-$921 per episode) for BPCI Model 3.

Reconciliation payments were higher than anticipated because participants were not required to make payments in excess of their target amounts during a portion of the intervention period. Even if negative NPRA had been retrieved from participants during this period, we estimate that Medicare would have realized a loss of $55.2 million due to BPCI Model 3.

IV. Impact of BPCI on Total Market Volume of Non-Fracture MJRLE Discharges

The BPCI initiative has the potential to incentivize providers to generate additional episodes. Under BPCI, incentives to increase volume go beyond those in the traditional Fee-for-Service (FFS) payment system because of the opportunity to achieve positive NPRA. Further, if the generated episodes were initiated for patients who could have foregone the procedure or service, then the episodes are likely to be less intensive than average, further bolstering opportunities to generate NPRA. If volume of discharges increases in response to BPCI, Medicare savings would decline by the amount of the payments for these additional episodes. BPCI participants may increase their volume or market share of particular clinical episodes through enhanced marketing, higher quality, or new gainsharing agreements with referring physicians. Shifts in patient volume across providers, however, would not have as much of an effect on Medicare savings.[84] In this analysis, we estimate the impact of the BPCI initiative (all EI types in Models 2, 3, and 4) on the volume of non-fracture MJRLE discharges in a market. We focused the analysis on the volume of non-fracture MJRLE discharges since, as elective procedures, providers have more influence over patient decisions to have the surgery, and because across all 48 clinical episodes in BPCI, MJRLE had the largest episode volume and the greatest number of participating providers.

A. Key Findings

We found that, on average, BPCI did not have a statistically significant impact on the volume of non-fracture MJRLE discharges within health care markets.[85] Our point estimates indicate that BPCI resulted in a non-statistically significant average increase in non-fracture MJRLE discharges of 0.71% for markets with a BPCI presence and a non-statistically significant average increase of 2.14% for BPCI providers. However, those point estimates are not statistically different from zero, even at a 20% significance level. Therefore, we conclude that BPCI had no effect on the volume of non-fracture MJRLE discharges for markets or providers.

B. Methods

We analyzed the impact of the BPCI initiative on the volume of MJRLE discharges in a market by testing whether markets (Medicare CBSAs (MCBSAs)) with a higher BPCI "dose" experienced larger or smaller increases in the MJRLE discharge rate (discharges per 1,000 Medicare FFS beneficiaries) than they would have otherwise. We measure BPCI "dose" as the market share of BPCI participants (hospitals, PGPs, SNFs, HHAs, IRFs, and LTCHs) prior to the start of the BPCI initiative. Almost all markets (87%) had some BPCI market share, but market shares varied widely across markets (see **Appendix D**). We examined volume at the market level to measure net increases in volume rather than shifts between non-BPCI providers and BPCI providers or among BPCI participants.

[84] The effect on Medicare savings of shifts in volume across providers would depend on the difference in episode payments between the providers.

[85] We also estimated the impact of BPCI on the volume of fracture MJRLE discharges and found no statistically significant impact (p-values for three terms that measure BPCI impact, which are defined in the next section, are 0.20, 0.70, and 0.64). This result was expected since fracture MJRLE procedures are not planned surgeries.

To account for the varying stages of the implementation of the BPCI initiative, we define three intervention periods:

(1) **Post 1 (Q4 2012 – Q3 2013)** is the intervention time period in which no MJRLE Awardees were in the risk-bearing phase of BPCI,[86]

(2) **Post 2 (Q3 2013 – Q3 2015)** is the intervention time period in which some MJRLE Awardees were in the risk-bearing phase of BPCI, some had not yet joined BPCI, and some had terminated participation in the MJRLE clinical episode, and

(3) **Post 3 (Q4 2015 – Q2 2017)** is the intervention time period in which all MJRLE Awardees were either in the risk-bearing phase of BPCI or had terminated participation in the MJRLE clinical episode.

We interact each of the "post" variables with the baseline BPCI market shares to measure each market's exposure to BPCI activity in the intervention time periods. We use a market-level DiD model using Ordinary Least Squares (OLS) and regress the discharge rate (discharges per 1,000 Medicare FFS beneficiaries) in a given market and quarter on market-level characteristics, market and quarterly fixed effects, and the three post intervention period indicators that are interacted with baseline market shares. In examining Medicare claims data to prepare for this analysis, we found that the number of non-fracture MJRLE discharges has increased across the nation for almost two decades. There is also a large degree of variation in the underlying trends in non-fracture MRJLE discharges across markets. Because of this, we include a MCBSA-specific linear time trend in the model. Our regressions are weighted by the number of Medicare FFS beneficiaries in the MCBSA because our interest is in the population discharge rate. [87,88]

We use the point estimates from our DiD OLS regression to predict the discharge rates attributable to BPCI and compare them to a counterfactual prediction in which we assume no BPCI initiative. This provides an estimate of the percentage change in market volume due to BPCI.[89] We also multiply the market percentage change by the ratio of total discharges to BPCI discharges in the market to provide an estimate of the percentage change in the volume of BPCI providers. The methodology for constructing these estimates is provided in **Appendix D**.

C. Results

The estimated coefficients and p-values are listed in Exhibits 77A and 77B. Our point estimates indicate that BPCI resulted in a non-statistically significant average increase in non-fracture MJRLE discharges of 0.71% for markets with a BPCI presence and a non-statistically significant average increase of 2.14% for BPCI providers. None of the DiD coefficients are statistically

[86] We consider this period an early intervention time period because it reflects the time after the Initiative was announced, but prior to when Awardees began to participate. During this time, potential applicants were analyzing their opportunity under BPCI and receiving data from CMS to evaluate that opportunity. In January 2013, CMS announced the organizations selected to participate in Phase 1 (non-risk bearing phase) of BPCI Models 2-4.

[87] Angrist, Joshua D., and Jörn- Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press.

[88] Card, David, and Alan B. Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." Journal of Political Economy 100(1):1–40.

[89] We do not calculate standard errors or confidence intervals for these percent changes.

significant and all have high p-values. On that basis, we conclude that BPCI had no statistically significant impact on non-fracture MJRLE volume for markets or BPCI providers.

Exhibits 77A and 77B: Estimates of the impact of BPCI on the volume of non-fracture MJRLE discharges

Variables Measuring BPCI Impact	DiD	P-Value
BPCI_MktShr X Post1	0.033	0.54
BPCI_MktShr X Post2	0.086	0.26
BPCI_MktShr X Post3	0.077	0.48

Unit of analysis	Percent change in volume
Market	0.71%
BPCI providers*	2.14%

Note: This analysis includes 401 MCBSAs; R-squared equals 0.952. We expect markets to have differential impacts of BPCI on volume. The difference-in-differences (DiD) estimates represent an overall weighted average effect on volume due to BPCI. Results differ when MCBSA-specific trends are not included in the model, which underscores the need to control for these trends.

* The provider-level impact is calculated from the market-level DiD estimates and the FFS population in the market (see **Appendix D** for more information).

A limitation of our analysis is the use of MCBSAs as markets, which may not accurately reflect markets for the diverse set of providers – hospitals, PGPs, and PACs – in the model. However, there is no universally agreed upon definition for health care market. Secondly, because there are so few markets with zero BPCI market share, there is no strong comparison group at the market level. Lastly, our measurement of "BPCI dose" does not vary based on the duration of BPCI participation within the market. In other words, a market would have the same BPCI dose whether the BPCI participants dropped out of BPCI after one year or continued to participate in BPCI through the end of the initiative in 2018. However, it was necessary to construct the measure in that way so that the BPCI dose was not endogenous to performance under the BPCI initiative.

V. Discussion and Conclusion

A. Discussion

This fifth annual BPCI evaluation report presents results based on an average of six quarters of experience for both Model 2 and Model 3 participants. Limited sample size made it difficult to draw conclusions about the impact of Model 4 on key outcomes; thus we do not present an evaluation of Model 4 in this report.[90] We estimated the impact of BPCI on select outcomes for 43 clinical episodes under Model 2 and 14 clinical episodes under Model 3. We continue to see general patterns of reduced intensity of PAC, with reductions in institutional care and increases in home health care. There are few indications in claims or beneficiary survey data that BPCI affected quality of care or functional status, either positively or negatively, although Model 2 BPCI patients had a slightly lower probability of reporting favorable care experiences and satisfaction.

With additional experience under the initiative, results remain consistent that BPCI participants responded to the initiative's incentives by reducing Medicare payments. This conclusion is strengthened because the results remain robust even after accounting for the significant number of episodes that were initiated by participants that ultimately stopped participating in that clinical episode. Although non-random withdrawal could have positively biased the payment estimates, the results remain when all episodes, even from participants after they stopped participating in the clinical episode, are included. Further, there is no evidence that the changes in utilization due to BPCI had disproportionate negative effects on vulnerable beneficiaries, as defined by dementia, dual eligibility, or prior institutional care.

Because the initiative was voluntary, it needed to be designed to encourage providers and other entities to participate. This was done primarily through reconciliation payments that participants could earn by lowering their episode payments. In addition, participants could choose among multiple options, including clinical episodes, which allowed them to customize the initiative to fit their particular circumstances. While these flexibilities influenced the participation in the initiative and also provided variability that contributed to understanding the features of bundling that were associated with reduced payments, they may also have resulted in the lack of savings to the Medicare program. If downside risk had not been eliminated from October 2013 through December 2014 for Model 2 hospitals and Model 3 SNFs and HHAs, and October 2013 through September 2016 for Model 2 PGPs, Model 2 would have resulted in $144.3 million savings to the Medicare program and Model 3 loses would have been cut by approximately $30 million. Other features of the initiative, particularly with respect to the determination of the target price, including the lack of risk adjustment and that it was based on participant-specific historical payments, likely also contributed to the inability to achieve Medicare program savings. Again, changes to the target price methodology to achieve savings would probably have affected participation.

B. Limitations

The primary analytic approach for this evaluation is dependent on how well the comparison group represents what would have happened absent the BPCI initiative. An unbiased DiD estimate

[90] See the CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 3 Evaluation & Monitoring Annual Report for details about Model 4. Available for download at https://innovation.cms.gov/initiatives/bundled-payments/

requires a matched comparison group that is similar to BPCI providers on key factors expected to influence their decision to participate in BPCI. In addition, because the DiD estimate attributes differences in trends between BPCI and the comparison group during the intervention period to the BPCI initiative, it is essential that the two groups have parallel trends for a given outcome during the baseline period. With these goals in mind, we matched providers and episodes on several factors, including payment and quality outcomes. In most Model, episode initiator type, clinical episode combinations, the comparison group represented a close match to the BPCI providers on these outcomes. For some combinations, the comparison episodes were not as close a match as we would like, even after multiple attempts to improve the match. In some cases, we rejected the null hypothesis that there were parallel trends for key quality and total payment impact estimates tested; we rejected 63 of 497 results, or 13% ($p<0.10$).[91] Thus, for these estimates, the underlying assumptions of the DiD method were violated, which may bias our results. In some instances, even when we failed to reject the parallel trend hypothesis, there were large differences in baseline levels of the outcome, which raises questions about whether the BPCI and matched comparison group had the same underlying trend in that outcome. For example, a high readmission rate among BPCI episodes in the baseline sample due to an extreme value could lead to a large difference in average baseline readmission rates between BPCI and comparison episodes. In this example, we would expect the differences in readmission rates to narrow during the intervention period, even in the absence of BPCI, as the estimated average in the BPCI intervention sample converges to the long-term average rate.

The majority of the analyses in this report are risk-adjusted to account for differences in provider and market characteristics, as well as patient mix that is measurable with claims data. We conducted sensitivity tests of most of the analyses, for example including patient characteristics from patient assessments, but as with all regression models, it is possible that we did not control for all characteristics that may affect the outcomes.

As a result of the limitations summarized above, our results for some individual outcomes among specific Model, episode initiator type, and clinical episode combinations may be biased. However, our overall conclusions that bundled payments has reduced episode payments while maintaining quality of care remains due to the consistency over time, across outcomes, clinical episodes, and robustness checks.

Our approach to creating the evaluation sample resulted in the exclusion of some BPCI and comparison episodes from the impact estimates. In order to assess the impact this approach had on our estimates, we conducted sensitivity analyses on key payment and quality outcomes that included the unmatched BPCI and comparison episodes. Instead of relying on a one-to-one matching between BPCI and comparison episodes, the sensitivity analysis uses all episodes from BPCI and comparison providers and weights comparison episodes by the number of BPCI providers the comparison provider matches. The results from the sensitivity analysis suggested that the statistical significance of some of the impact estimates may be due to the random sample of comparison episodes that were included in the matched sample, and they may not have been statistically significant if a different sample of episodes had been selected. However, these

[91] We could not conduct the test for an additional 25 DiD estimates due to small sample size. Because we tested the null hypothesis that there were parallel trends at the 10% significance level, this proportion is slightly above the 10% that would be observed by chance alone.

sensitivity tests have their own limitations. The comparison weights do not take into account that BPCI providers are matched with different numbers of comparison providers. BPCI participants that are matched to a higher number of comparison providers are overrepresented in the comparison sample, which may create a bias if they are different from BPCI participants that are matched to a smaller number of comparison providers. In addition, we cannot determine statistical significance of the sensitivity test estimates because the standard errors are inaccurate, as the standard error calculations treat the data as if the weighted episodes represent multiple independent data points when they do not.

With respect to the survey results, because we do not have survey data predating the BPCI initiative, we cannot be certain whether different responses between patients treated by BPCI and comparison providers were caused by the BPCI initiative or existed prior to BPCI. In addition, the BPCI PGP beneficiary survey sample included some episodes that may not have been assigned to participating PGPs for purposes of reconciliation because the BPCI PGPs did not provide any post-discharge care for these beneficiaries.[92] Preliminary estimates suggest that over one-third of episodes fall in this category. This could result in an underestimate of the association between BPCI and survey outcomes.

In addition, we cannot disentangle the exact reasons for statistically significant relative declines in payments for the ITT DiD estimates. For example, there may not have been enough episodes from providers after they withdrew to shift the estimate, or providers may have continued with care redesign after withdrawing, resulting in continued reductions in Medicare payments relative to the comparison group. Similarly, we cannot determine the exact reason some ITT DiD estimates were not statistically significant, though it is likely that BPCI providers did not reduce payments relative to the comparison group at the same rate when no longer faced with BPCI incentives as they did during Phase 2.

The analysis of factors associated with variation in NPRA among BPCI providers is correlational. Therefore, we were unable to determine if the observed associations between NPRA and explanatory factors were due to BPCI. However, results from this analysis align with findings from the BPCI impact estimates. Specifically, this analysis found that top performing BPCI providers were associated with relative reductions in institutional PAC use. At the same time, we found that BPCI participation, particularly for hospital EIs, often resulted in a reduction in institutional PAC use compared to changes among similar providers not participating in BPCI, and this was associated with relative reductions in total payments.

The estimate of Medicare program savings required several assumptions. First, we extrapolated the impact of BPCI on payments in the analysis sample to all clinical episodes, including some for which we did not produce BPCI impact estimates. Thus we assumed the analysis sample was representative of all clinical episodes. Second, we used BPCI intervention episodes from the analytical sample to calculate a conversion rate from standardized Medicare payments to non-standardized payments and assumed the ratio of non-standardized to standardized Medicare

[92] In order to be a PGP episode included in reconciliation, the beneficiary must have at least one post-discharge visit during the episode with a clinician billing under the PGP's TIN. In order to survey the beneficiary within 90 days of hospital admission, we did not have the claims necessary to identify which beneficiaries met this additional requirement.

payments was similar for the baseline and intervention period, as well as for BPCI and comparison episodes. Third, we assumed no change in episode volume.

C. Conclusion

The evidence is mounting that bundled payments can reduce payments for multiple clinical episodes without compromising quality of care. After considering the NPRA that CMS paid to participants under the BPCI design, Medicare spending was higher under the initiative relative to what would have been spent. However, if downside risk had not been eliminated, Model 2 may have resulted in savings, and Model 3 losses would have been much less.

The promising results of Model 2 have informed the development of BPCI Advanced, which will qualify as an Advanced Alternative Payment Model. Among the many differences between BPCI and BPCI Advanced are several key changes designed to help ensure net Medicare savings. BPCI Advanced features modified target prices that incorporate risk adjustment and reflect peer performance and a higher discount. Changes to the target prices are intended to encourage both high and low cost providers to participate, which would lessen the self-selection we have seen in BPCI. Some BPCI clinical episodes were not included in BPCI Advanced due to high clinical heterogeneity or small volume. In addition, the participant entry and exit opportunities are scaled back under BPCI Advanced. BPCI Advanced will start at the conclusion of BPCI and will be implemented under the authority of CMMI. Like all models tested by CMS, there will be a formal, independent evaluation to assess the impact of BPCI Advanced, including changes in quality of care and Medicare savings as well as any unintended consequences.

There is still more to learn from the BPCI initiative. There are seven more quarters of claims and assessment data to evaluate. The additional data will allow more in-depth analyses of particular participant types, market effects, and beneficiary sub-populations, for example. Additional research is needed to calibrate what type and level of financial rewards are required to entice participants into reducing episode payments without completely offsetting those lower payments.