Project Evaluation Activity in Support of Partnership for Patients

Interim Evaluation Report, Final

September 2015

Revised 12/29/2015

Centers for Medicare and Medicaid Innovation
Contract No. GS-10F-0166R
Task Order: HHSM-500-2014-00440G

HSAG HEALTH SERVICES ADVISORY GROUP
MATHEMATICA Policy Research

Partnership for Patients: Interim Evaluation Report, Final

Contract Number: GS-10F-0166R
Task Order: HHSM-500-2014-00440G

Submitted To:
Kevin L. Frazier, MSc, MPA
Technical Advisor
Centers for Medicare & Medicaid Services
Centers for Clinical Standards & Quality
Quality Improvement Innovation Models Testing Group
7500 Security Boulevard
Woodlawn, MD 21244

September 2015

Revised 12/29/2015

Health Services Advisory Group
3133 East Camelback Road, Suite 300
Phoenix, AZ 85016-4501
Program Director: Tim Laios

Mathematica Policy Research
600 Maryland Avenue, SW, Suite 550
Washington, DC 20024-2512
Mathematica Project Director: Suzanne Felt-Lisk

Health Services Advisory Group
Robert Fornango, PhD
Gina M. Tignini
Cindy Strickland, JD
Rochelle Malinoff, PhD, MBA
Michael Lichter, PhD
Ravee Nithianandam, MS
Nikhil Nagaraj, MS
Bin Gu, MS
Jason Martindale
Sunny Bateman
Julia Lewis
Malissa Mojica
Brian Starr

Mathematica Policy Research*
Valerie Cheh, PhD
Suzanne Felt-Lisk, MPA
Jelena Zurovac, PhD
Keith Kranker, PhD
Jessica Heeringa, MPH
Lori Timmins, PhD
Andrea Wysocki, PhD
Javier Rodriguez, PhD
Linda Barterian, MPP
Mariel Finucane, PhD
Arnold Chen, MD, MSc
Catherine McLaughlin, PhD
Kirsten Barrett, PhD
Tyler Fisher
Lauren Vollmer, MA, MS
Alex Bryce, MS
Boriana Pratt, MA, MS

*Lead for this report

Contents

Executive Summary

The Partnership for Patients (PfP) campaign was launched in April 2011 with the ambitious goals of reducing preventable hospital-acquired conditions (HACs) by 40 percent and 30-day hospital readmissions by 20 percent. At that time, the Agency for Healthcare Research and Quality (AHRQ) reported 14.5 patient harms were occurring per 100 discharges.[1] To reduce harm at the desired level of magnitude, the campaign originally selected nine focus areas of harm, as well as readmissions, for which there was evidence that improvement was possible, with a tenth focus area, obstetrical early elective delivery (OB-EED), being added in early 2012.[2]

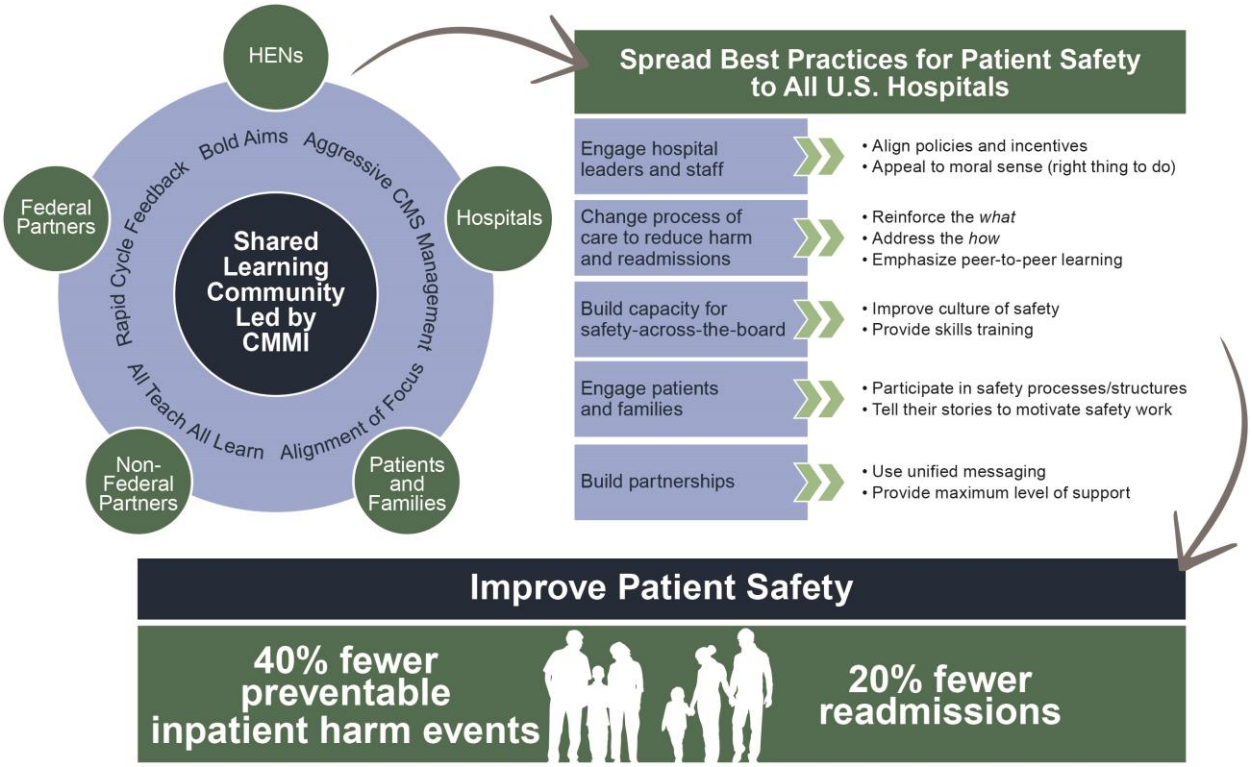
Other public and private stakeholders focused on financial incentives and improvement work for some of these areas, but the high level of harm that persisted more than a decade after the landmark 1999 Institute of Medicine (IOM) report, *To Err is Human*, suggested that the pace of harm reduction had been slow despite knowledge of the problem. PfP set bold aims, then implemented a strategy to align these other healthcare stakeholders—including federal and other public and private healthcare payers, providers, and patients—to build partnerships that focused on reducing harm and accelerating improvement. With relevant stakeholders moving in the same direction at the same time, the campaign strove to overcome the inherently limited reach of any single initiative operating in the complex healthcare environment.

Central to the campaign was the investment of considerable Center for Medicare & Medicaid Innovation (CMMI) resources to create a learning community, which sought to spread best practices by engaging hospital executives and staff, supporting changes in process of care, building the capacity of hospitals to improve patient safety, and engaging patients and families (Figure 1). The community (and campaign) was national in scale, and CMMI strove to have an impact on all hospitals in the United States (U.S.). From the population of all hospitals, 72 percent of general acute care hospitals in the U.S., representing over 80 percent of admissions, worked with one aspect of the campaign, Hospital Engagement Networks (HENs), during 2012-2014. HENs were tasked with recruiting hospitals to work with the PfP campaign, collaborating with other partners and contractors, providing feedback to their hospitals, and helping to facilitate activity within the learning communities. The campaign was managed aggressively and continuously, by the Centers for Medicare & Medicaid Services (CMS), to constantly challenge hospitals and other stakeholders to do more to achieve the goals.

1 AHRQ. Interim Update on 2013 Annual Hospital-Acquired Condition Rate and Estimates of Cost Savings and Deaths Averted From 2010 to 2013. Available at http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/pfp/interimhacrate2013.pdf. Accessed on August 29, 2015.

2 The nine focus areas are as follows: adverse drug events (ADE), catheter-associated urinary tract infection (CAUTI), central line-associated blood stream infections (CLABSI), falls, other OB adverse events (OB-Other), pressure ulcers, surgical site infections (SSI), ventilator-associated pneumonia (VAP), and venous thromboembolism (VTE).

Figure 1—The PfP Campaign Strategy



Purpose of the Report

This is an interim evaluation report for the PfP campaign. The goals of the evaluation were to assess the reduction in inpatient harms and readmissions during the period 2011–2014, and to assess, if possible, the contribution that PfP made to those reductions. Table 1 shows characteristics of PfP (some of which are unusual in intervention/evaluation cycles) that collectively challenged the evaluation's ability to attribute change to the campaign. The national scale of PfP, coupled with the partnerships with other stakeholders targeting the same areas, are the biggest factors that made it challenging to assess the progress that would have been made in its absence. The Evaluation Contractor employed multiple data sets and methodologies—mixing qualitative and quantitative analyses—in an effort to understand more fully the effects of PfP on harm reduction. The implications of harm reduction for healthcare costs were also addressed, providing input to U.S. Department of Health and Human Services (HHS) in its assessment of whether the investment in PfP paid off for taxpayers.

Table 1—PfP Characteristics That Are Atypical and Challenged the Evaluation

Characteristic	PfP Characteristics Atypical for Intervention/Evaluation Cycles
Scale	All patients within acute care hospital inpatient settings nationally
Partnerships	Major strategic component
Concurrent Work with Other Initiatives Targeting Similar Objectives	Widespread and actively promoted for many of the focus areas
Definition of Intervention	Locally defined

Table 1—PfP Characteristics That Are Atypical and Challenged the Evaluation	
Characteristic	**PfP Characteristics Atypical for Intervention/Evaluation Cycles**
Change to Participation During Funded Period	Dynamic–continual entry and exit
Measurement Philosophy	Allow local choice in measures and flexibility in baselines and reporting periods to encourage widespread participation

Overview of Data and Methods

Data sources used in this evaluation included secondary data (for example, Medicare claims, a database of medical chart abstractions known as the Medicare Patient Safety Monitoring System [MPSMS], and outcomes data submitted by HENs), survey data (two hospital surveys—a national survey to a random sample of hospitals and a survey targeted to HEN-aligned hospitals), and qualitative data including interviews with HENs and site visits to a dozen hospitals. Statistical process control (SPC) charts and interrupted time series (ITS) methods were used to assess trends over time. Although the PfP campaign was deliberately national in scope and attempted to impact all hospitals nationwide, approximately 28 percent of hospitals did not officially align with the PfP campaign. Had the PfP campaign succeeded in receiving commitments from 100 percent of hospitals nationally, there would be no opportunity to compare outcomes in HEN-aligned hospitals to those of non-HEN-aligned hospitals.[3] However, because nearly 30 percent of hospitals did not officially align with the campaign, the Evaluation Contractor had the opportunity to construct a comparison group and assess whether officially aligning with a HEN produced a greater reduction in patient harms than hospitals that did not formally align with a HEN.[4] Therefore, difference-in-differences regression analysis was used to estimate the impacts of the HEN component of PfP by comparing changes in outcomes of patients in HEN-aligned hospitals with those in a comparison group of hospitals, using propensity score reweighting to make the groups statistically equivalent on observed characteristics. Both frequentist and Bayesian difference-in-differences analyses were used. There are limitations to each data source and each analysis; the results from any one analysis, whether based on qualitative or quantitative data and methods, cannot stand alone as an evaluation of the PfP campaign or the HEN component of that campaign (see Chapter 1). The evaluation was designed to enable triangulation of the results stemming from these various data sets and methodologies, in hopes that through consideration of all the qualitative and quantitative results together, that a clearer picture of the PfP impact on patient harm would emerge.

[3] HENs were tasked with recruiting hospitals to participate in the PfP campaign by committing to reducing patient harms, reporting data to HENs, participating with the HEN and other contractors, and actively working toward achieving the PfP goals. Hospitals that officially committed to working with a HEN in the PfP campaign are referred to as HEN-aligned. HEN-aligned hospitals received collaboration, support, and feedback from the HEN with which they were aligned (see chapter 3 for further details). While non-HEN-aligned hospitals may have also worked with HENs and other contractors in an unofficial capacity, the extent to which this happened is unknown. However, the extent of support and feedback received from the HENs was likely to be less than that received by HEN-aligned hospitals.

[4] The HENs participating in the PfP campaign were encouraged to open their activities, workshops, and collaborative workgroups to non-HEN-aligned hospitals as well as those that were officially aligned. To the extent that the openness of the campaign was successful in reaching and influencing hospitals that did not officially commit to the campaign, the magnitude of any difference in reduction in patient harms between HEN-aligned and non-HEN-aligned hospitals may be reduced.

Possible Outcomes

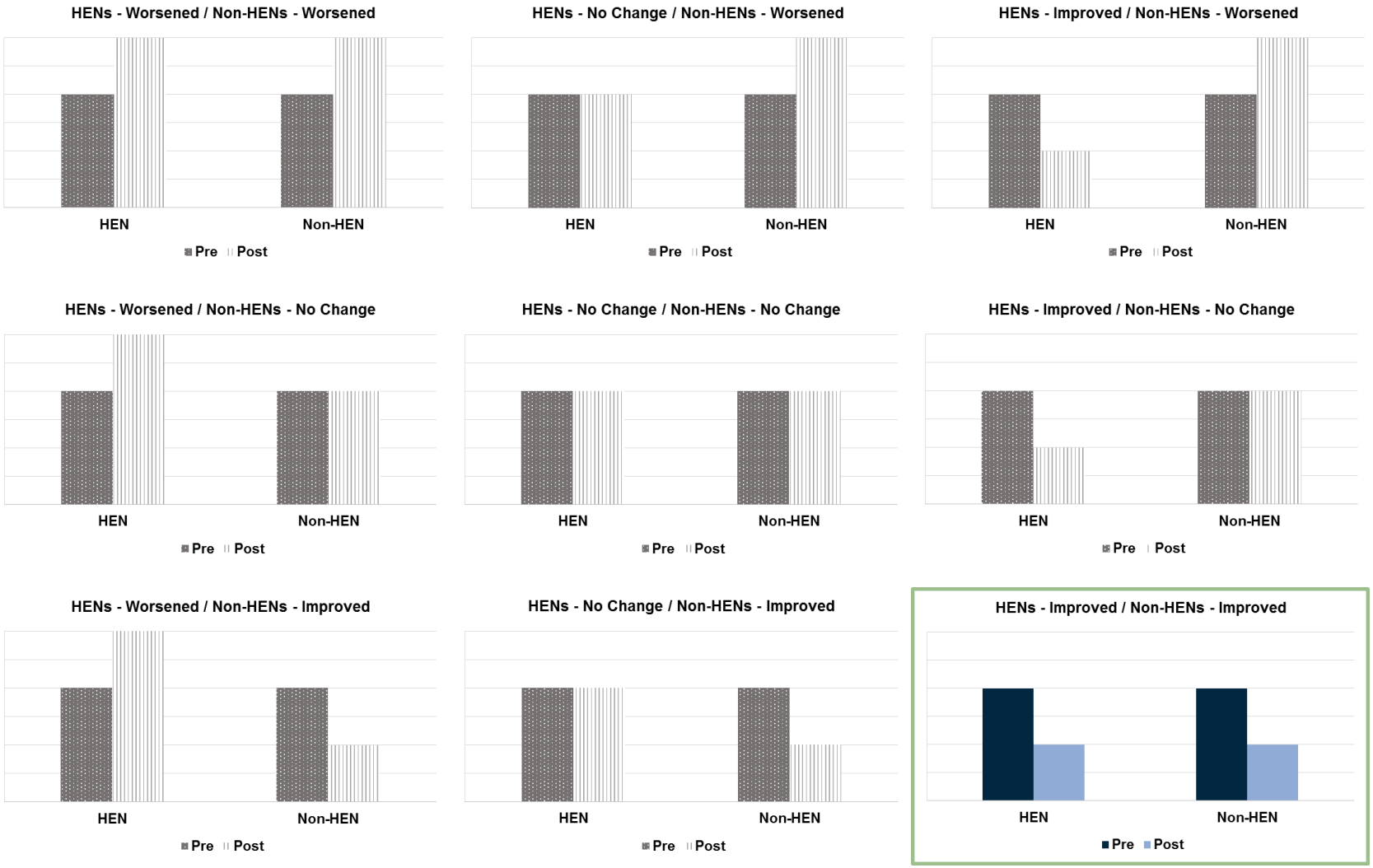
To place the results comparing the HEN-aligned and non-HEN-aligned hospitals in the PfP campaign into context, it is useful to present and explain several of the outcomes that could have been observed during the period of the campaign, including the outcome that actually occurred. Patient harms among both HEN-aligned and non-HEN-aligned hospitals could have improved, worsened, or exhibited no change, forming a 3-by-3 matrix of possibilities. Figure 2 presents a conceptual description of the nine potential outcomes with respect to whether HEN-aligned and non-HEN-aligned hospitals improved, worsened, or exhibited no change in rates of patient harms.[5]

In the difference-in-difference analyses, a clear impact of the campaign is identified when the campaign participants do better relative to those who did not participate. For example, HEN-aligned hospitals could have exhibited improvement or no change in patient harms, while non-HEN-aligned hospitals worsened. Alternatively, HEN-aligned hospitals could have exhibited improvements in patient harms, while non-HEN-aligned hospitals exhibited no change in patient harms. These scenarios are shown in the panels on the upper right portion of Figure 2. In contrast, the lower left portion of the figure contains panels in which the non-HEN-aligned hospitals could have outperformed HEN-align hospitals, which would therefore call into question the efficacy of the campaign.

The lower right panel of Figure 2 is highlighted in green to illustrate the results that were observed most often in the difference-in-differences analyses of HEN-aligned and non-HEN-aligned hospitals. In this panel, both groups of hospitals often exhibited significant reductions in patient harms; however, there was no difference in the magnitude of the change that the Evaluation Contractor could detect. Of the nine possible types of outcomes, the outcome that was observed most often is the most favorable–that is, harms decreased for patients at both the HEN-aligned and non-HEN-aligned hospitals. Depending on the harm category and analysis, the results exhibit some variation, and will be highlighted throughout the report where appropriate.

[5] The conceptual diagram in Figure 2 represents the possible outcomes relative to the direction of changes for patient harms and not the magnitude of change. As such, the Evaluation Contractor acknowledges that other possible outcomes could have been observed. For example, had HEN-aligned hospitals worsened but less so than non-HEN-aligned hospitals, this could have been considered evidence of a programmatic impact on patient harms. Ignoring the magnitude of changes, the nine panels in Figure 2 are sufficient to describe the possible outcomes for the PfP campaign in a conceptual manner.

Figure 2— Nine Results That Were Possible in the PfP Campaign



Note: The set of outcomes actually observed for the aggregate rate of harm from the AHRQ National Scorecard was that both HEN-aligned and non-HEN-aligned hospitals improved. See panel outlined in green on the bottom right side of the figure.

Key Findings

Overall, national rates of inpatient harm and Medicare fee-for-service (FFS) readmissions have markedly improved since the start of the campaign. A substantial portion of the data used in this report covers the period through CY 2013, with a smaller number of data sources and measures extending into 2014. Since 2010, inpatient harm shows reductions at a national level consistent with the PfP goal of reducing inpatient harm by 17.6 percent by the end of 2014.[6] These reductions, as shown in Table 2, include over 5,000 fewer venous thromboembolism (VTE) events, over 50,000 fewer falls, over 118,000 fewer OB-EEDs, and 45,000 fewer surgical site infections (SSI) (also see Chapter 6). Readmissions dropped as well, with an estimated 222,000 readmissions avoided, although this reduction remained well short of the 20 percent reduction goal.[7] For many harm areas, improvement began prior to PfP. This is not surprising since those who designed PfP chose to focus on harm areas for which there was already evidence that improvement was feasible, and other efforts to reduce these harms were underway. Additional analyses are ongoing through other HHS agencies using more recently available data. Figure 3 shows decreases in harm based on the AHRQ National Scorecard data, which is largely based on medical chart reviews; other sources and measures largely corroborated the overall pattern of decrease in harms (see Chapter 2).

Table 2–Estimate of Events Averted Due to National Decreases in Harm

PfP Harm Area	Reduction in Adverse Events During 2011-2013, Using 2010 Baseline (Rounded)
Adverse Drug Events (ADE)	577,000
Catheter-Associated Urinary Tract Infection (CAUTI)	190,000
Central Line-Associated Blood Stream Infection (CLABSI)	10,800
Falls	50,000
Obstetrical (OB)-Adverse Event	10,000
OB-Early Elective Delivery (OB-EED)[a]	118,000
Pressure Ulcers	280,000
Surgical Site Infections (SSI)	45,000
Ventilator-Associated Pneumonia (VAP)	8,000
Venous Thromboembolism (VTE)	5,000
All Other Hospital-Acquired Conditions (HACs)	142,000
Readmissions[b]	222,000
Total	**1,657,800**

6 This is the operational interpretation of 40 percent of preventable inpatient harm, based on the AHRQ estimate that 44 percent of inpatient harm may be preventable, and 40 percent of that equals 17.6 percent. Using data through 2014, AHRQ reported a reduction of 17.0 percent in patient harms and 2.1 million harms averted. The AHRQ report is available at http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2014.html,

7 Medicare FFS readmissions dropped 5.6 percent from 2010 through 2014 (see Chapter 2); non-Medicare data were available only through 2013 and show no decrease to that point (data provided by AHRQ).

Table 2–Estimate of Events Averted Due to National Decreases in Harm	
PfP Harm Area	**Reduction in Adverse Events During 2011-2013, Using 2010 Baseline (Rounded)**

Source: AHRQ publications available at http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html, http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2013.html, and readmissions estimates provided by AHRQ.
[a]OB-EED data uses 2010 baseline, events averted 2011–2013. Data from the NVSS (vital records data) for most U.S. counties.
[b]Readmission data use a 2010 baseline, with 2011 – 2013 estimate based on 1) Nationwide Readmissions Database (NRD) developed by AHRQ from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases; and 2) Medicare Provider Analysis and Review (MedPAR) files from CMS. Data provided by AHRQ staff member Joanna Jing, August 27, 2015.

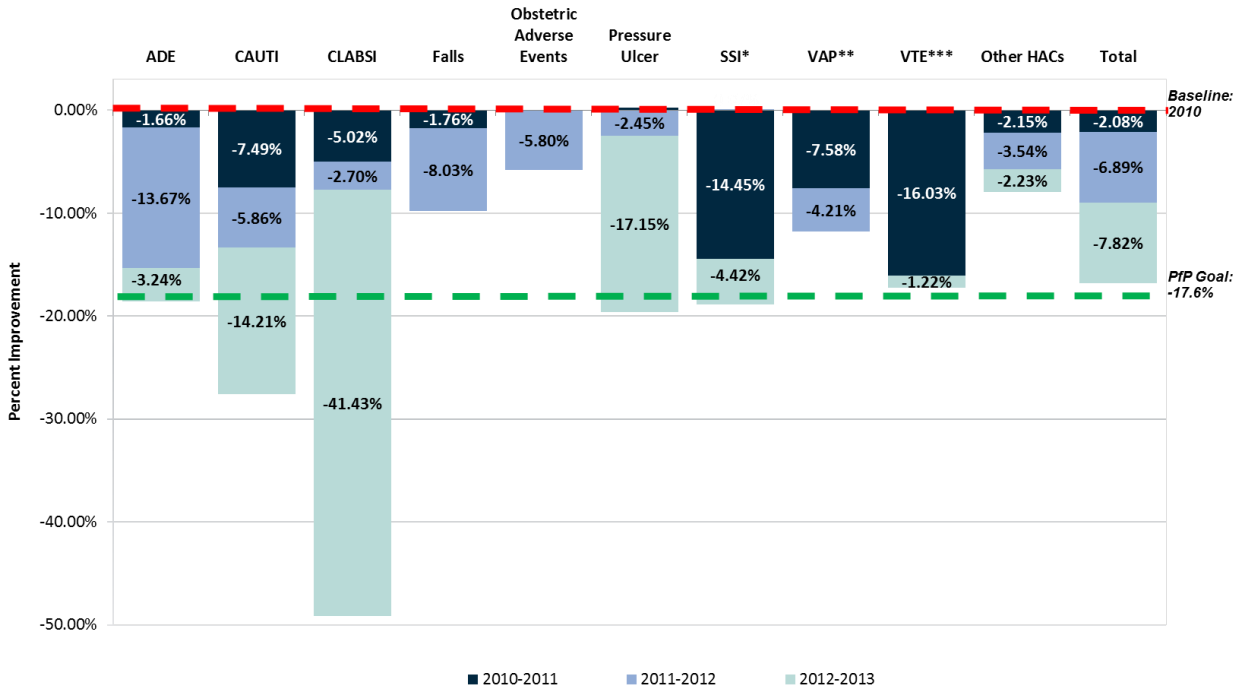
In the figures below, the Evaluation Contractor presents examples of the patient harm reductions observed in each harm area using a variety of the data sources available for the analyses. The charts presented here were selected because the rates were obtained from reliable data sources, provide good coverage of the populations of interest, and exhibit a representative degree of harm reduction observed on a national basis. For a complete listing of the harm reductions observed in all of the national measure trends, see Chapter 2. Importantly, due to differences in measure specifications, the results observed in data from some sources are not consistent with the results observed in the AHRQ National Scorecard data.

The national trend data were examined below, and in Chapter 2, using two methods: SPC charts and ITS regression analysis. The SPC chart results provide an assessment of the nature and timing of any special cause variations in the data series. Special cause variation refers to fluctuations that cannot be considered random variability over time, such as data points that fall outside of the empirically derived control limits to random variability, and shifts in the center line or average measure rate (see Appendix D for the detailed methodology for SPC charts).

ITS analysis was also performed using the data from national measures, to assess whether or not the series exhibit significant structural breaks (that is, a statistically significant shift in the average level of the series or change in the slope of the trend over time) after either the official PfP baseline period (i.e., Q1 2011), or the beginning of HEN operations (i.e., Q1 2012).[8] Analyses were performed to assess the significance of structural breaks associated with each of these periods. For some measures, however, the data series examined did not contain enough data prior to Q1 2011 or Q1 2012 to evaluate whether structural breaks occurred at these points. For measures where this is the case, it is not possible to ascertain whether observed trends in the data are the result of significant structural breaks associated with the timing of the PfP campaign, or if observed trends are the result of pre-existing secular trends in the data.

[8] The official baseline of the PfP campaign was calendar year 2010. Therefore, ITS analyses examined structural breaks in the first period following 2010, or Q1 2011. The HEN contracts were awarded on December 9, 2011, and represent the implementation of a major structural component of the PfP campaign. Thus, the Evaluation Contractor also used Q1 2012 as the first full quarter during HEN operations to assess structural breaks associated with the HEN component of the campaign.

Figure 3—Harm Reduction Estimated by AHRQ's National Scorecard, by Type of Harm and in Total, in Relation to PfP Goal



Rates (per 1,000 Discharges)	ADE	CAUTI	CLABSI	Falls	Obstetric Adverse Events	Pressure Ulcer	SSI	VAP	VTE	Other HACs	Total
Baseline 2010	49.50	12.21	0.55	7.94	2.50	40.31	2.93	1.16	0.85	27.30	145.25
Current 2013	40.30	8.85	0.28	7.20	2.36	32.51	2.38	1.12	0.71	25.14	120.86
Percent Change (2010-2013)	-18.57	-27.56	-49.15	-9.25	-5.63	-19.33	-18.81	-3.56	-17.25	-7.91	-16.80

Source: PfP's "AHRQ National Scorecard" or National HAC rate provided by Noel Eldridge on November 20, 2014, AHRQ Center for Quality Improvement and Patient Safety (CQuIPS).

Note: AHRQ's data cover 9 of the 10 PfP focus areas: adverse drug events (ADE), catheter-associated urinary tract infection (CAUTI), central line-associated blood stream infections (CLABSI), falls, OB-Other, pressure ulcers, SSI, ventilator-associated pneumonia (VAP), VTE, and "all other HACs." (Readmissions are not included.)

*There was an increase of 0.07 percent from 2011 to 2012 in SSI, after a -14.45 percent decrease from 2010 to 2011. The increase from 2011 to 2012 is not represented in the figure.

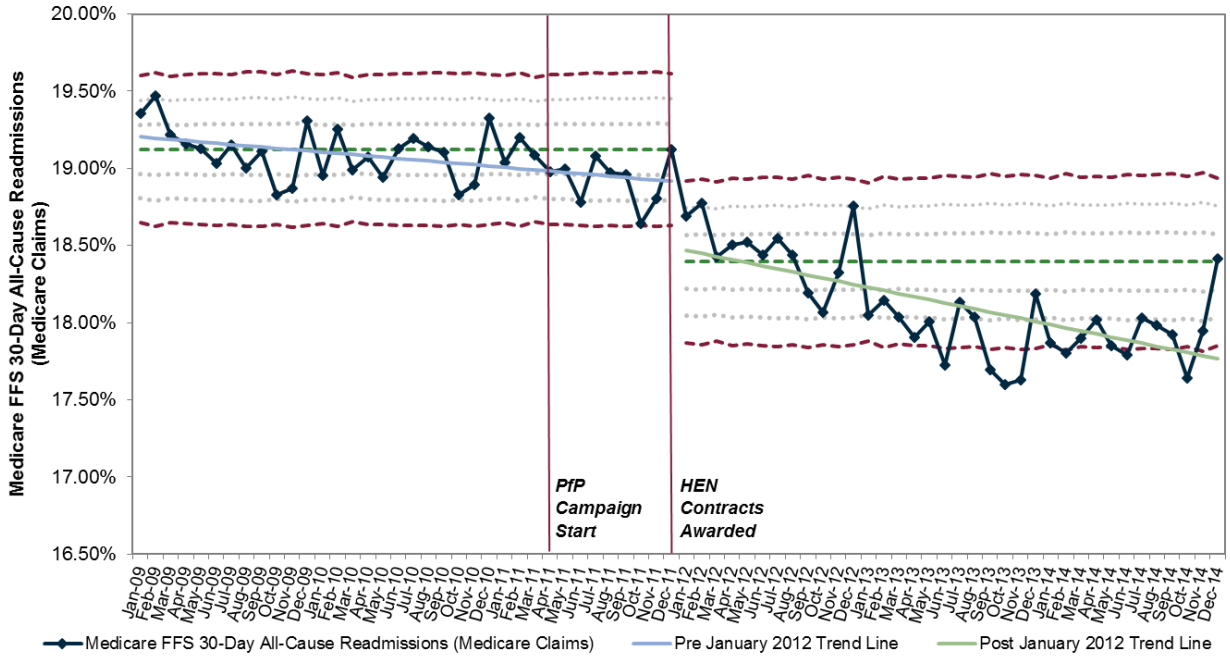
**There was an increase of 8.23 percent from 2012 to 2013 in VAP events after the -11.79 decrease from 2010 to 2012. The total decrease from 2010 to 2013 was 3.56 percent. The increase from 2012 to 2013 is not represented in the figure.

***There was an increase of 32.00 percent from 2011 to 2012 in VTE events after the -16.03 percent decrease from 2010 to 2011. In 2012-2013 there was a sharp -33.22 percent decline in VTE events which resulted in the total of a -17.25 percent decline from 2010-2013. The increase from 2011-2012 is not represented in the figure.

Readmissions

The SPC chart for the Medicare FFS 30-Day All-Cause Readmissions rate exhibits a shift in center line in January 2012 (Figure 4). A shift of this magnitude implies that the process (i.e., the real-world forces) generating these patient harms had changed sufficiently to result in a lower average rate. The ITS analysis of the series did not detect a significant structural break in January 2011. However, the readmissions rate exhibited a significant downward trend over time. In contrast, the ITS analysis detected a significant structural break in January 2012.

Figure 4—FFS 30-Day All-Cause Readmissions (Medicare Claims)

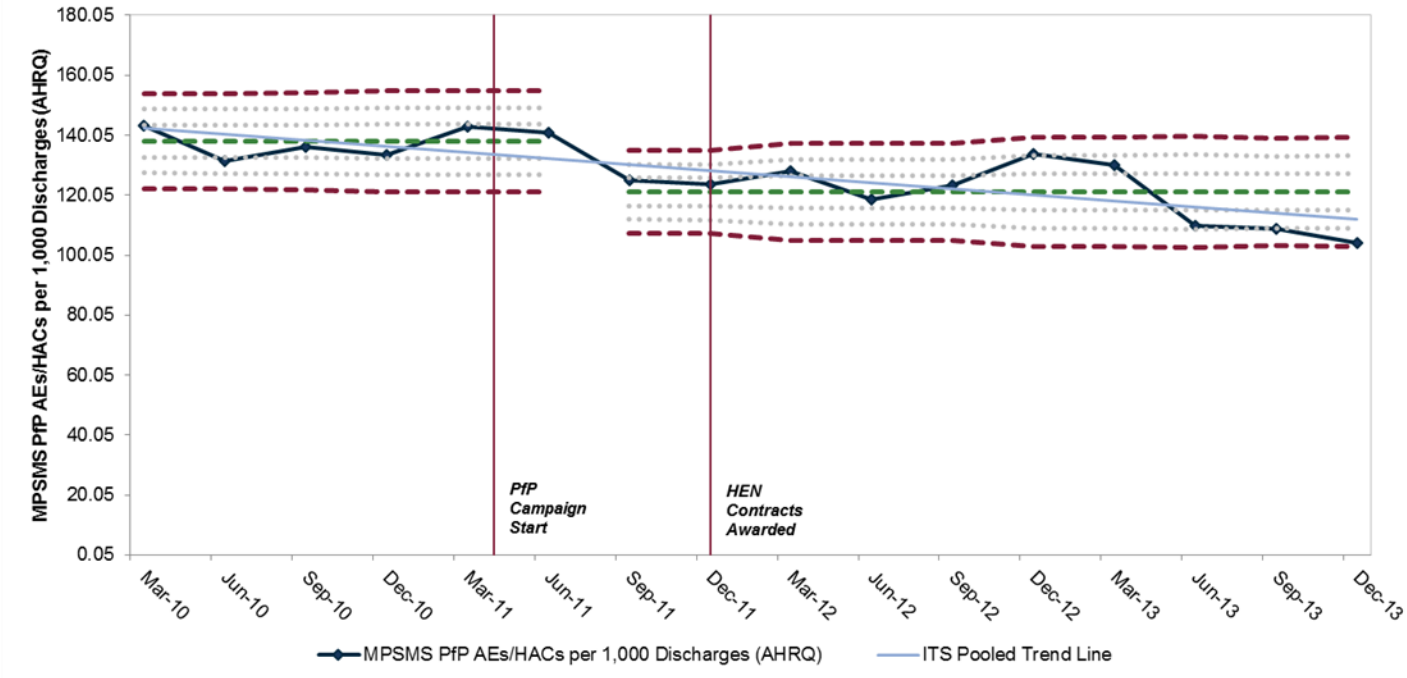


Source: Medicare claims data provided by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS. The Evaluation Contractor processed and ran regression-adjusted analysis to control for changing demographics independently, with similar findings.
Note: Center line and control limits (U' chart) for the first phase were calculated with data between January 2009 and March 2010. Center line and control limits (U' chart) for the second phase were calculated with data between January 2012 and March 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Data include between 981,065 and 754,486 discharges per month.

All Harm

The AHRQ National Scorecard data were trended quarterly and examined using SPC charts to identify evidence of special cause variation. ITS models were also estimated to test whether or not a significant structural break occurred in Q1 2011, the first quarter following the official 2010 PfP baseline. Figure 5 presents the quarterly AHRQ National Scorecard HACs rate per 1,000 discharges with the results of the SPC and ITS analysis. The SPC results indicate that a downward shift in the center line of the series was observed beginning in Q3 2011. The ITS analysis did not detect a significant structural break in Q1 2011, indicating that a single downward linear trend was sufficient to characterize the series. This result indicates that the downward trend in the 2010 AHRQ National Scorecard HACs rate continued without a statistically significant change through the 2011 to 2013 period. Finally, the ITS analyses are sensitive to the timing of the break point tested, and future analyses by other HHS agencies will use different analytic methods and alternative specifications to elaborate on these results.

Figure 5—AHRQ National Scorecard HACs per 1,000 Discharges (AHRQ)

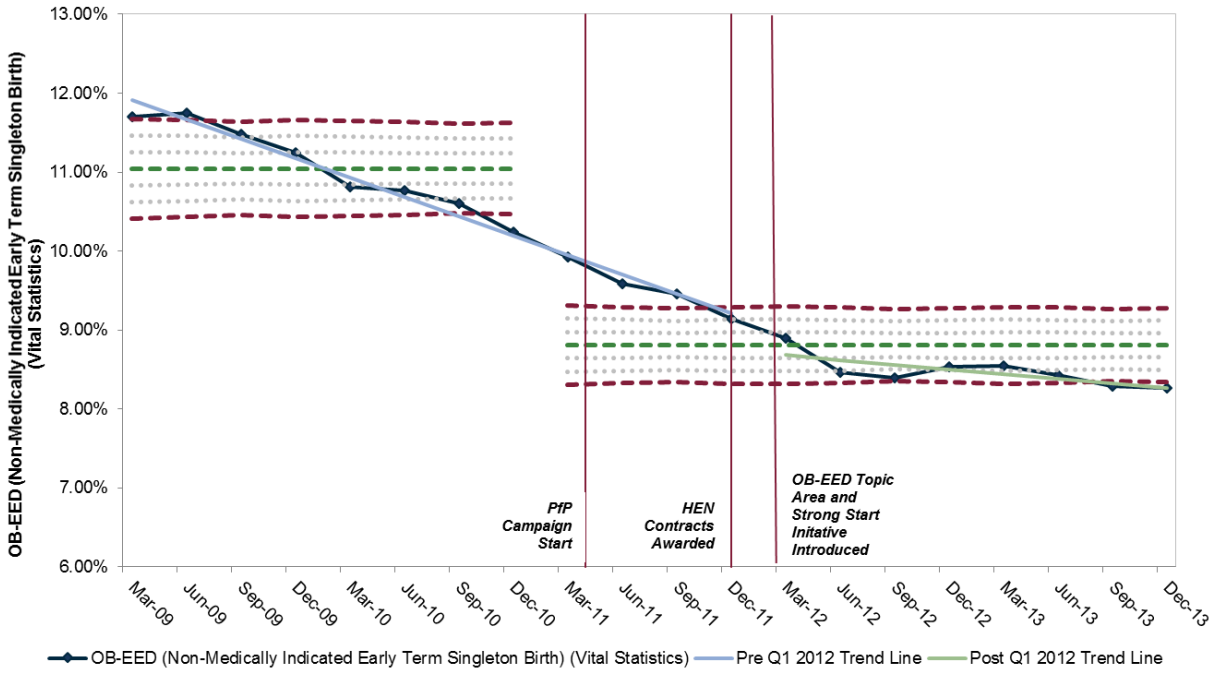


Source: AHRQ National Scorecard: 28 measures derived from MPSMS (chart-abstracted), AHRQ's Patient Safety Indicators (PSIs) (claims), and National Healthcare Safety Network (NHSN) (standardized hospital-reported to the Centers for Disease Control and Prevention [CDC]).
Note: Center line and control limits (U' chart) for first phase were calculated with data between Q1 2010 and Q2 2011. Center line and control limits (U' chart) for second phase were calculated with data between Q3 2011 and Q4 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

OB-EED

In the area of OB-EED, Vital Statistics data were examined using SPC and ITS analyses (Figure 6). The SPC chart results indicate a downward shift in the center line of the series in Q1 2011. In contrast, the ITS analyses detected a significant structural break in the series in Q1 2012. No structural break was identified using a Q1 2011 break point. Overall, the series exhibits a steady and strong decline in OB-EEDs from 2009 through 2011, with the rate of decline leveling off in 2012. However, the PfP campaign added OB-EED as a focus area in March of 2012, prior to the leveling off of the data series. Thus, some of the decline in OB-EED occurred during the time period of the PfP campaign, and it is possible that the campaign could have contributed to the decline.

Figure 6—OB-EED (Non-Medically Indicated Early Term Singleton Birth) (Vital Statistics)

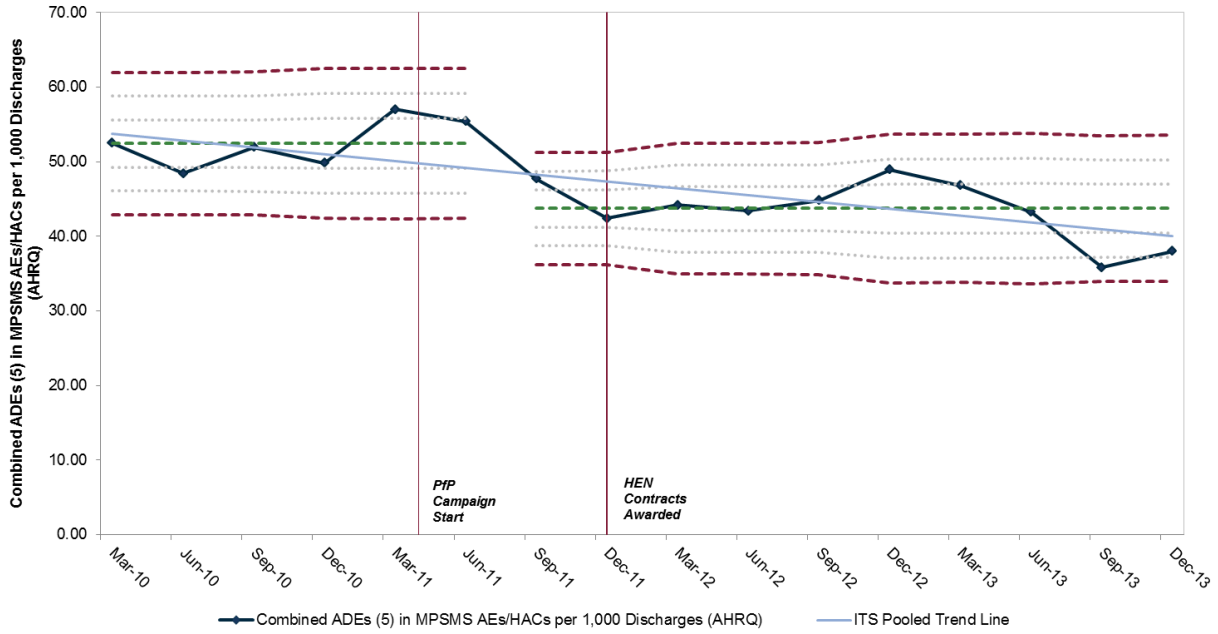


Source: National Vital Statistics System (NVSS).
Note: Center line and control limits (U' chart) for first phase were calculated with data between Q1 2009 and Q4 2010. Center line and control limits (U' chart) for second phase were calculated with data between Q1 2011 and Q4 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

ADE

Using the AHRQ National Scorecard data on ADE, the SPC chart observed a slight downward shift in the center line of the series in Q3 2011 (Figure 7). However, the ITS analyses did not detect a structural break in the series in Q1 2011 or Q1 2012. Thus, a linear trend beginning in 2010 continued unabated through the end of 2013. The series examined here does not contain data from 2014, and therefore cannot identify any reductions in ADE that may have occurred during the last year of the PfP campaign.

Figure 7—Combined ADEs (5) in MPSMS Adverse Events/HACs per 1,000 Discharges (AHRQ)



Source: AHRQ National Scorecard (quarterly data).
Note: Center line and control limits (U' chart) for first phase were calculated with data between Q1 2010 and Q2 2011. Center line and control limits (U chart) for second phase were calculated with data between Q3 2011 and Q4 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

VTE

Medicare claims data used to calculate Patient Safety Indicator (PSI)-12 measuring perioperative pulmonary embolism (PE) or deep vein thrombosis (DVT) illustrate a downward trend in patient harm rates from Q2 2011 through Q4 2014 (Figure 8). The SPC chart also observed a downward shift in the center line for the series in Q3 2013. In contrast, the ITS analysis did not detect a structural break in the linear trend beginning in Q1 2012. No structural break for Q1 2011 could be tested due to a lack of data prior to this period.

Figure 8—Perioperative PE or DVT per 1,000 Surgical Discharges (Medicare Claims)

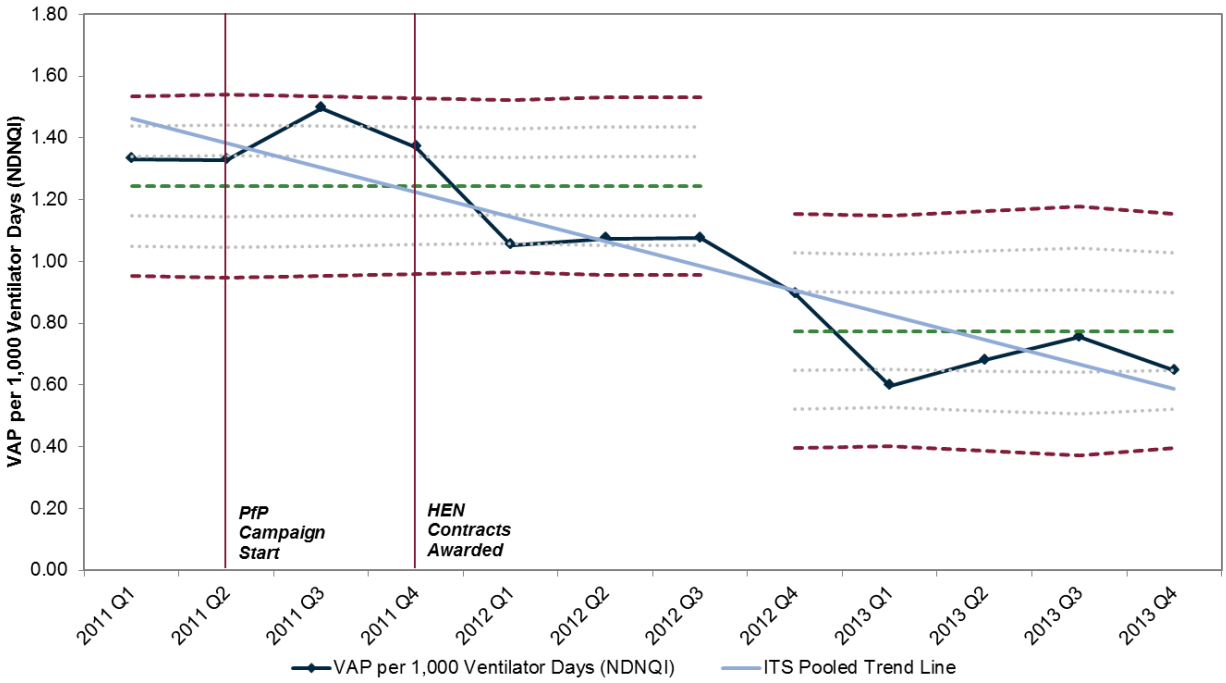
Source: Rates calculated by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS based on Medicare FFS claims data. The Evaluation Contractor also conducted a similar analysis, adding a longer time series (from 2009) and using regression analysis to control for changing demographics over time. The data showed a similar pattern.

Note: Control limits and center line (U′ chart) for the first phase were calculated with data between Q2 2011 to Q1 2013.Control limits and center line (U' chart) for the second phase were calculated with data between Q2 2013 to Q4 2014.The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma lines; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicare FFS claims for all hospitals that reliably report present on admission (POA) status (≥ 95 percent of the hospital's diagnoses for a given quarter are accompanied by a valid code for POA) and that have the following characteristics: all hospitals paid under Medicare's inpatient prospective payment system (IPPS), critical access hospital (CAH), cancer hospitals, and Maryland hospitals. Data include between 619,182 and 690,101 discharges per quarter.

Ventilator-Associated Pneumonia (VAP)/Ventilator-Associated Event (VAE)

The VAP rate per 1,000 ventilator days also exhibits a downward trend between Q1 2011 and Q4 2013 among the approximately 500 hospitals reporting to the National Database of Nursing Quality Indicators® (NDNQI®), a voluntary database maintained by the American Nurses Association (ANA) (Figure 9). The SPC chart observed a downward shift in the center line in Q4 2012. In contrast, the ITS analysis did not detect a structural break in the series in Q1 2012. No structural break for Q1 2011 could be tested due to a lack of data prior to this period.

Figure 9—VAP per 1,000 Ventilator Days (NDNQI)[9]



Source: NDNQI. Data are between 484 and 544 hospitals per quarter.
Note: Control limits and center line (U' chart) for first phase constructed using data from Q1 2011 to Q3 2012. Control limits and center line (U' chart) for second phase constructed using data from Q4 2012 to Q4 2013. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. VAP data after Q4 2013 are not comparable due to a change in definition and a large reduction in the number of reporting hospitals.

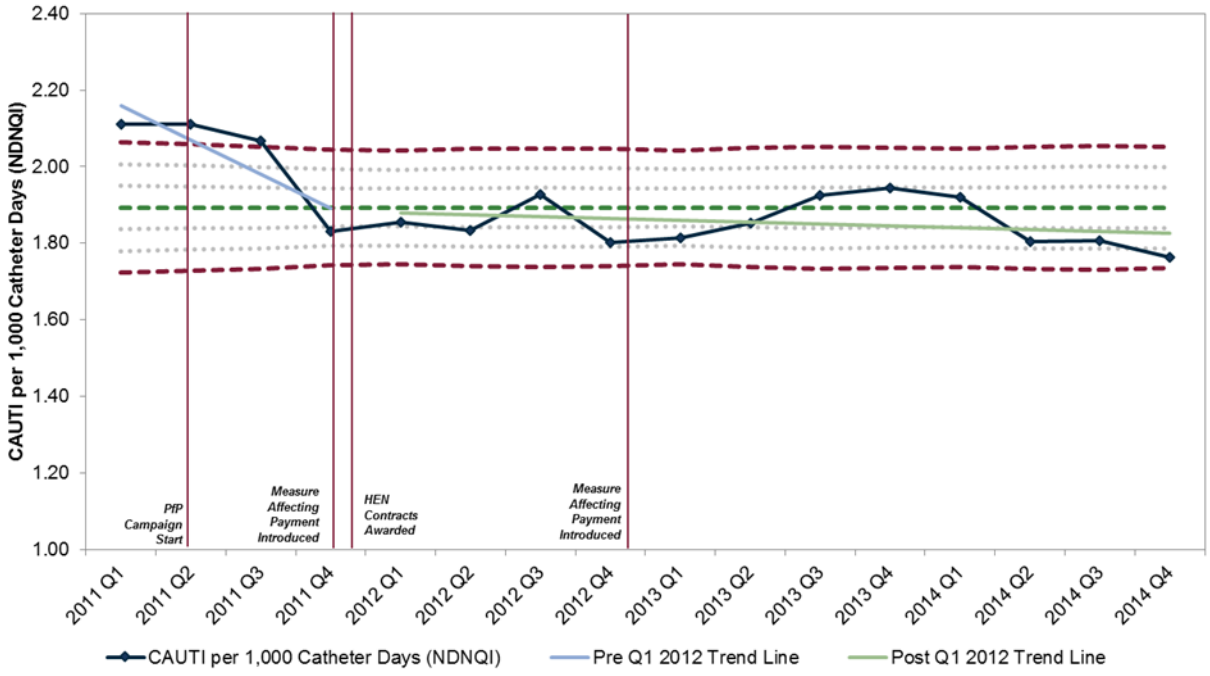
[9] NDNQI® is a registered trademark of the ANA. NDNQI® data were supplied by ANA. The ANA disclaims responsibility for any analyses, interpretations, or conclusions.

Catheter-Associated Urinary Tract Infection (CAUTI)

The CAUTI rate per 1,000 catheter days exhibited a slight decline during 2011 among the approximately 600 hospitals reporting to the NDNQI, a voluntary database maintained by the ANA (Figure 10). The SPC chart shows three points above the upper control limits between Q1 2011 and Q3 2011, although there was no other evidence of special cause variation. The ITS analyses detected a structural break in Q1 2012, when the declining trend from 2011 slowed appreciably to a nearly flat trend. No structural break for Q1 2011 could be tested due to a lack of data prior to this period.

Importantly, the CAUTI rate presented here differs in definition from that collected in the MPSMS data used for the AHRQ National Scorecard, which accounts for some of the difference in performance. However, as shown in Chapter 2, catheter utilization was also declining at the same time as the number of infections, causing both the numerator and denominators to decline, and contributing to the lack of a downward trend in the data.

Figure 10—CAUTI per 1,000 Catheter Days (NDNQI)

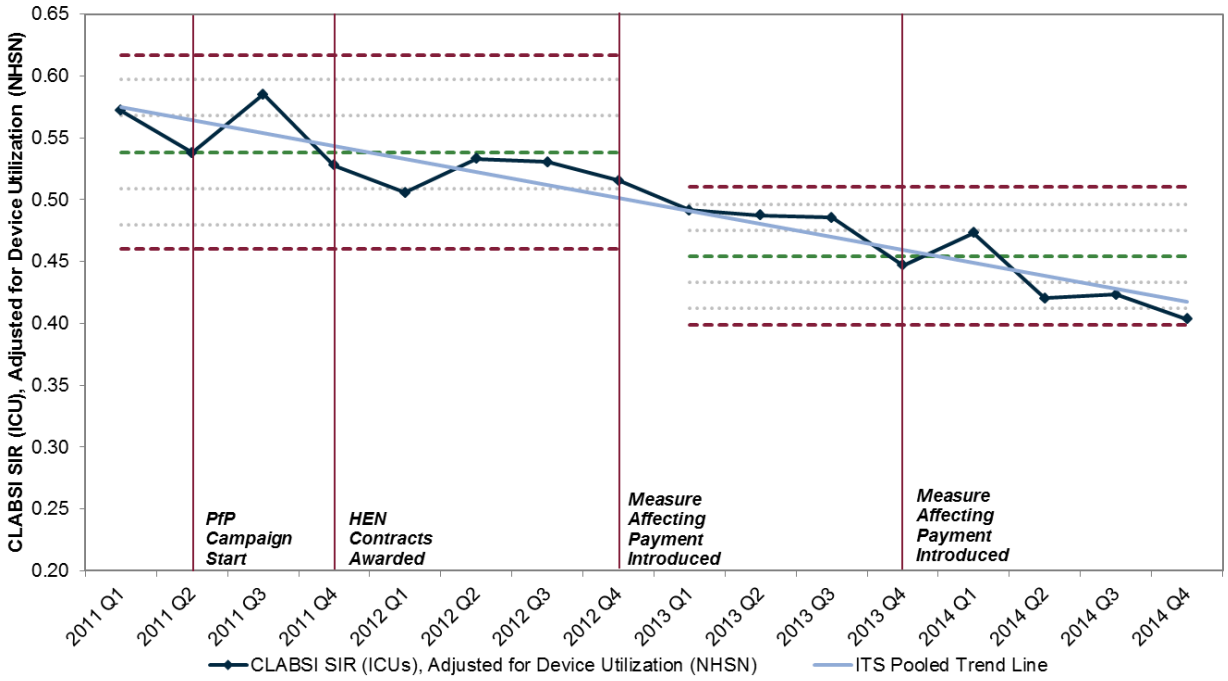


Source: NDNQI. Data are between 554 and 680 hospitals per quarter.
Note: Control limits and center line (U′ chart) were constructed using data from Q1 2011 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Central-Line Associated Blood Stream Infection (CLABSI)

The CLABSI standardized infection ratio (SIR) from the Centers for Disease Control and Prevention's (CDC's) National Healthcare Safety Network (NHSN) data declined steadily between Q1 2011 and Q4 2014 (Figure 11). The SPC chart observed a downward shift in the center line for the series in Q1 2013. In contrast, the ITS analysis did not detect a significant structural break in the trend in Q1 2012. No structural break for Q1 2011 could be tested due to a lack of data prior to this period.

Figure 11—CLABSI SIR for ICUs, Adjusted for Device Utilization (NHSN)



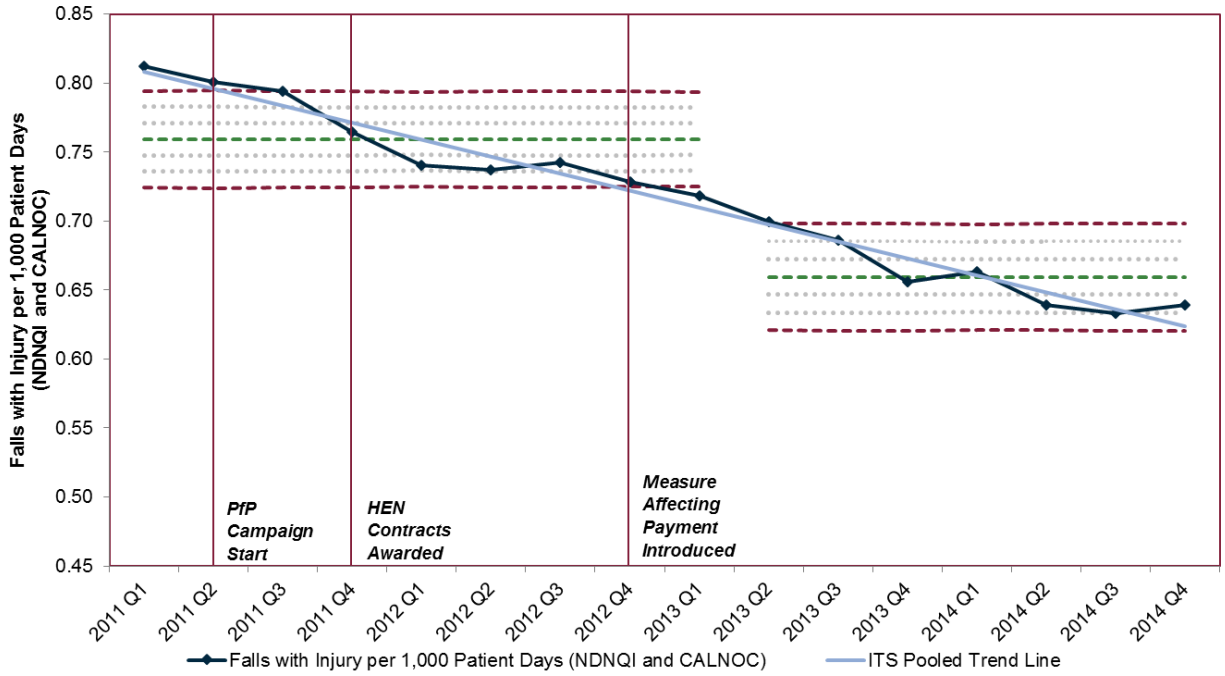
Source: NHSN. Data are between 3,048 and 3,152 hospitals per quarter. Data are adjusted for change in utilization, which also changes risk of harm. To adjust, the national unadjusted CLABSI SIR (ICUs) data points are multiplied by the ratio of the national device utilization rate for that period to the national device utilization rate in the first quarter.

Note: Center line and control limits (X chart) for the first phase were calculated with data between Q1 2011 and Q4 2012. Center line and control limits (X chart) for the second phase were calculated with data between Q1 2013 and Q4 2014. While the PfP campaign began in April 2011, HEN contracts were not awarded until December 2011; only modest educational support occurred in 2011. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. The data from NHSN were not provided with clustered standard errors that would account for the non-independence of discharges within the same facilities. Without the clustered standard errors, the SPC charts and ITS analysis performed are less conservative than would be the case if the correlation of outcomes within facilities were accounted for in the analysis.

Falls

The falls with injury per 1,000 patient days measure from approximately 1,300 hospitals voluntarily reporting to NDNQI and the Collaborative Alliance for Nursing Outcomes (CALNOC) also exhibited a steady, linear decline between Q1 2011 and Q4 2014 (Figure 12). The SPC chart observed a downward shift in the center line of the series in Q2 2013. In contrast, the ITS analysis did not detect a significant structural break in Q1 2012. No structural break for Q1 2011 could be tested due to a lack of data prior to this period.

Figure 12—Falls with Injury per 1,000 Patient Days (NDNQI and CALNOC)



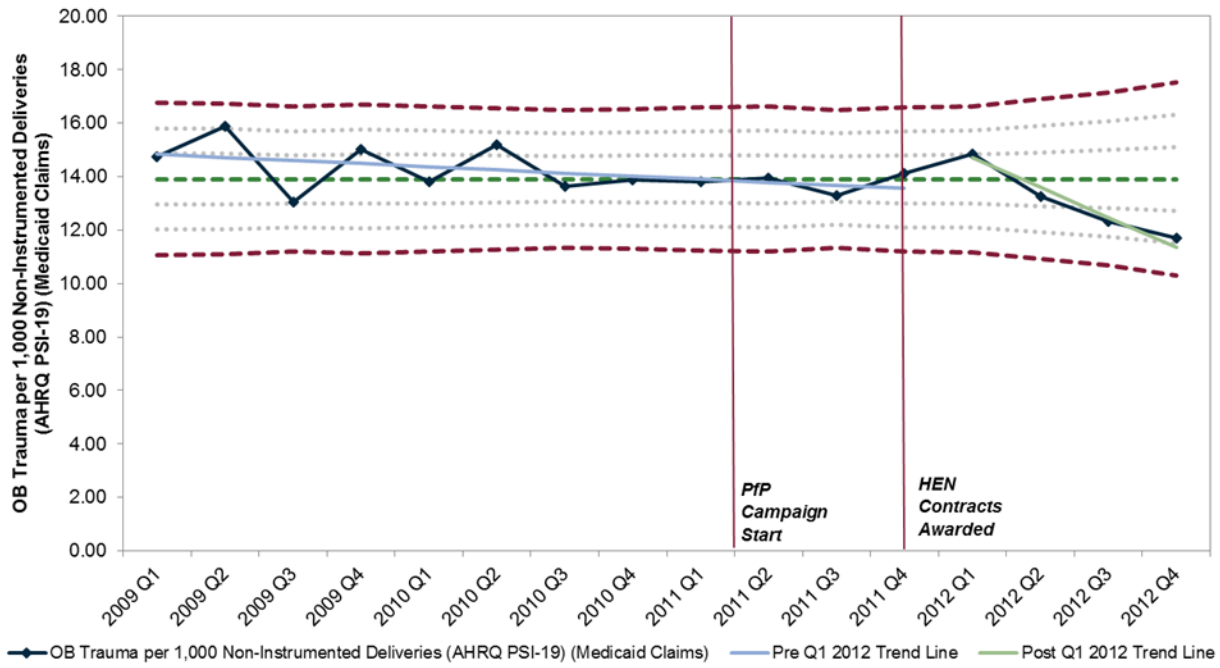
Source: NDNQI and CALNOC. Data are between 1,268 and 1,342 hospitals per quarter.

Note: Control limits and center line (U′ chart) for period Q1 2011 to Q1 2013 were constructed using data from Q1 2011 to Q1 2013. Control limits and center line (U′ chart) for period Q2 2013 to Q4 2014 were constructed using data from Q2 2013 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Other OB Adverse Events (OB-Other)

The obstetric trauma measure per 1,000 non-instrumented deliveries drawn from Medicaid claims for 17 states remained relatively flat between Q1 2009 and Q4 2012 (Figure 13). The SPC chart did not detect any evidence of special cause variation. However, the ITS analysis detected a significant structural break in Q1 2012, after the HEN efforts began. In contrast, no structural break was detected in Q1 2011.

Figure 13—Obstetric Trauma per 1,000 Non-Instrumented Deliveries (AHRQ PSI-19) (Medicaid Claims)

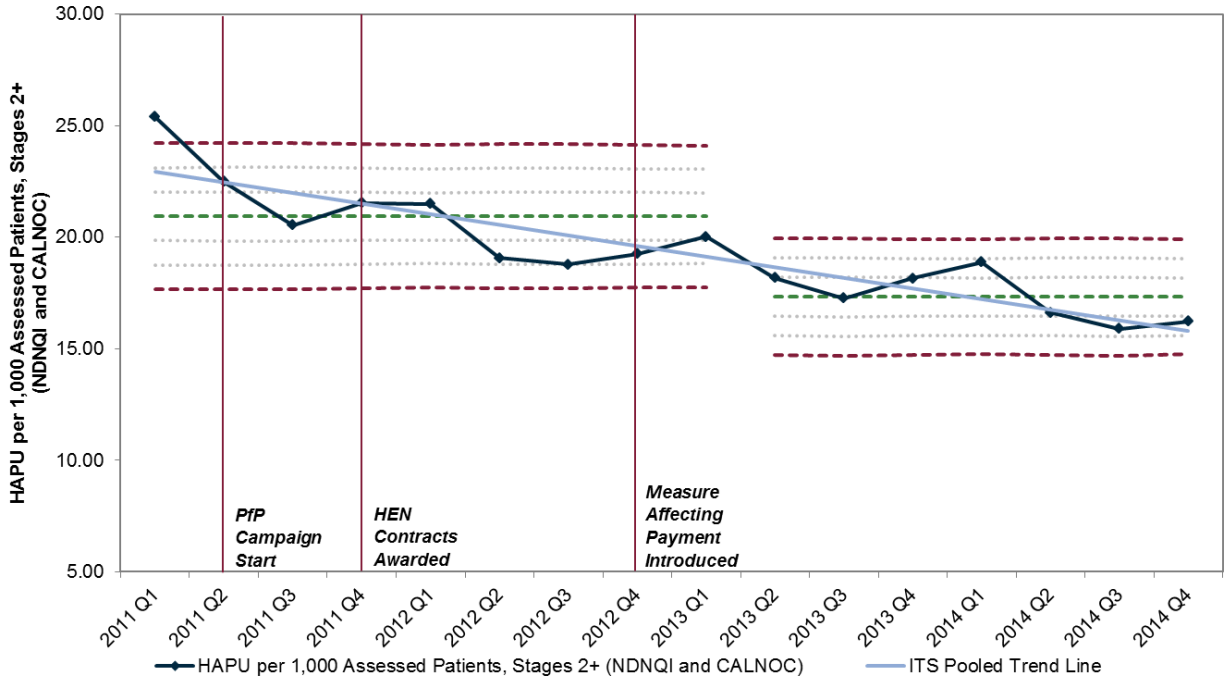


Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q4 2012. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, Medicaid Managed Care [MMC] and FFS data in 3 states), assuming all diagnosis codes were not POA (Medicaid claims do not include POA data) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 21,410 and 42,577 discharges per quarter.

Pressure Ulcers

The hospital-acquired pressure ulcer rate for stage 2 and higher for about 1,300 hospitals voluntarily reporting to NDNQI or CALNOC declined steadily between Q1 2011 and Q4 2014 (Figure 14). The SPC chart observed a downward shift in the center line of the series in Q2 2013. In contrast, the ITS analysis did not detect a significant structural break in Q1 2012. No structural break for Q1 2011 could be tested due to a lack of data prior to this period.

Figure 14—Hospital-Acquired Pressure Ulcer (Stages 2+) per 1,000 Assessed Patients (NDNQI and CALNOC)

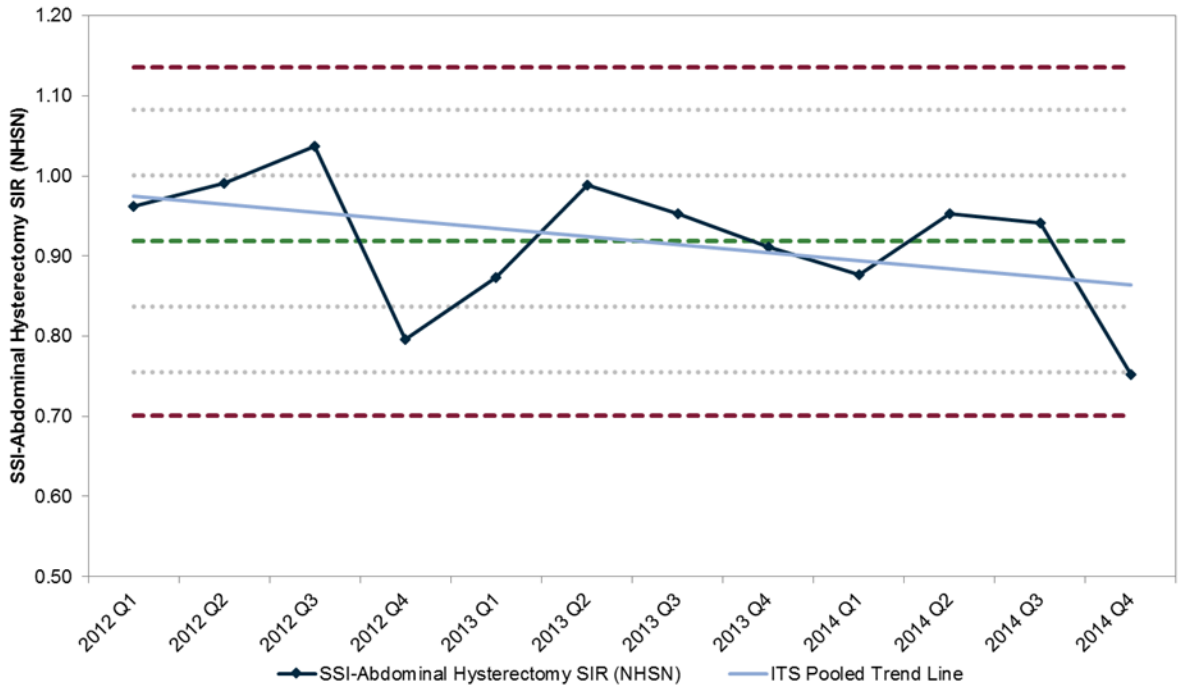


Source: NDNQI and CALNOC. Data are from between 1,291 and 1,357 hospitals per quarter.
Note: Control limits and center line (U′ chart) for period Q1 2011 to Q1 2013 were constructed using data from Q1 2011 to Q1 2013. Control limits and center line (U′ chart) for period Q2 2013 to Q4 2014 were constructed using data from Q2 2013 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

SSI

The SSI SIR for abdominal hysterectomy procedures declined between Q1 2012 and Q4 2014 (Figure 15). However, the SPC chart analysis did not detect any evidence of special cause variation. Additionally, the ITS analysis did not detect a significant structural break in the series in Q1 2013. No structural break for Q1 2011 or Q1 2012 could be tested due to a lack of data prior to this period. It is important to note that this measure focuses on SSIs for a single procedure, whereas the data in the AHRQ National Scorecard focus on 17 procedures.

Figure 15—SSI Abdominal Hysterectomy SIR (NHSN)



Source: NHSN. Data have 3,340 hospitals per quarter.
Note: Control limits and center line (X chart) constructed using data from Q1 2012 to Q4 2014. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma lines; and the dotted lines just inside the control limits are the two-sigma limits. The data from NHSN were not provided with clustered standard errors that would account for the non-independence of discharges within the same facilities. Without the clustered standard errors, the SPC charts and ITS analysis performed are less conservative than would be the case if the correlation of outcomes within facilities were accounted for in the analysis.

Table 3 summarizes changes in harm for each harm area based on all data sets. In addition to the AHRQ National Scorecard measures shown above, all other available data sets show a decrease in harm for ADE, readmissions, SSI, VAP, OB-EED, and falls. All available data sets except Medicaid claims—which were only available through 2012 or Q1 2013, unlike the other data which include at least through 2013—show decrease in harm for pressure ulcers, VTE, and CLABSI. CAUTI and OB harms other than OB-EED are the two areas where some measures in the Medicare or all-payer data show improvement and others do not. Whether the pace of improvement accelerated, remained steady, or slowed during the PfP campaign depends on the data sources and measures used, although for many metrics and sources, the pace remained steady (right-hand column). For the measures exhibiting no change in trend, the results imply that pre-existing trends in the patient harm rates continued after the PfP campaign began. For measures in which trend improved or worsened, the change in these measures coincides with the beginning of the PfP campaign.

Table 3—National Trend Results by PfP Harm Area

PfP Harm Area	National Trend in Rate of Harm[c] (Data Source Abbreviation)	ITS Analysis of National Trend (Change in *Trend–Q1 2011 Breakpoint*)
ADE	**Decrease** AHRQ	**Improved Trend** AHRQ
VTE	**Decrease** Medicare (Claims), AHRQ, HENs (Self-Reported) **No Change** Medicaid (Claims, 2012)	**No Change** Medicaid (Claims)
Pressure Ulcers	**Decrease** NDNQI/CALNOC, Medicare (Claims), AHRQ **Increase** Medicaid (Claims, 2012)	**Worsening Trend** Medicaid (Claims)
Readmissions	**Decrease** Medicare (Claims), Medicaid (Claims, Q1 2013), AHRQ-Readm, HENs (Self-Reported)	**No Change** Medicaid (Claims, Children Q1 2013) **Improved Trend** Medicare (Claims), Medicaid (Claims, Adult 2012)
CLABSI	**Decrease** NHSN (Self-Reported Using CDC Case Definition), AHRQ, NDNQI, Medicaid (Claims, 2012) **No Change** Medicaid (Claims, Q1 2013)	**Improved Trend** Medicaid (Claims, 2012, Q1 2013)
CAUTI	**Decrease** AHRQ, NDNQI, Medicaid (Claims, 2012[a]) **No Change** NHSN (Self-Reported Using CDC Case Definition) **Increase** Medicaid (Claims, 2012[a])	**No Change** Medicaid (Claims, 2012, Q1 2013[b]) **Worsening Trend** Medicaid (Claims, Q1 2013[b])

Table 3—National Trend Results by PfP Harm Area		
PfP Harm Area	**National Trend in Rate of Harm[c] (Data Source Abbreviation)**	**ITS Analysis of National Trend (Change in *Trend–Q1 2011 Breakpoint*)**
VAP	**Decrease** NDNQI, AHRQ	—
SSI	**Decrease** AHRQ, NHSN (Self-Reported Using CDC Case Definition)	—
OB-EED	**Decrease** NVSS, HENs (Self-Reported)	**No Change** NVSS
OB-Other	*OB Trauma* **Decrease** HENs (Self-Reported), Medicaid (Claims, 2012), AHRQ *Injury to Neonate* **No Change** HENs (Self-Reported) **Decrease** Medicaid (Claims, Q1 2013) *Assisted Ventilation* **Decrease** NVSS *APGAR Score 0-6* **No Change** NVSS *Admissions to NICU* **Increase** NVSS *Low Birth Weight* **No Change** NVSS	*OB Trauma* **No Change** Medicaid (Claims, 2012) *Injury to Neonate* **No Change** Medicaid (Claims, Q1 2013) *Assisted Ventilation* **No Change** NVSS *APGAR Score 0-6* **Worsening Trend** NVSS *Admissions to NICU* **Worsening Trend** NVSS *Low Birth Weight* **No Change** NVSS
Falls	**Decrease** AHRQ, NDNQI/CALNOC, Medicare (Claims)	**No Change** NDNQI/CALNOC, Medicare (Claims)
Aggregate Measures of Harm	**Decrease** AHRQ	**No Change** AHRQ

Table 3—National Trend Results by PfP Harm Area		
PfP Harm Area	**National Trend in Rate of Harm[c] (Data Source Abbreviation)**	**ITS Analysis of National Trend (Change in *Trend–Q1 2011 Breakpoint*)**

Source: Evaluation Contractor's analysis of data sources as listed.
Note: Data run through at least 2013 unless another end period is noted.
[a]In the Medicaid claims data, two measures were tested: the CAUTI HAC measure and a measure for Hospital-acquired UTIs (HAUTI)—which includes many more infections some of which are not related to catheters. For HAUTI, the pediatric rate decreased while the adult rate increased through 2012. For the CAUTI HAC, the reverse was true.
[b]In the Medicaid claims data, ITS analysis of change in trend using Q1 2011 as the structural break point showed no change for CAUTI –adults and HAUTI-pediatrics, and worsening for HAUTI-adults.
[c]"Decrease," "No Change," or "Increase" in the first column means a decrease was found through consideration of the preponderance of available evidence from ITS analyses, SPC chart evidence of special cause variation, or the visual trend showed the indicated direction.
Data source abbreviations (See Appendix B for a description of each source):
AHRQ = AHRQ National Scorecard, composed primarily of data from MPSMS obtained through chart abstraction.
AHRQ-Readm = All-payer readmissions rates calculated by AHRQ, provided through Noel Eldridge, personal communication, August 26, 2015.
Medicare = Medicare FFS claims.
Medicaid = Medicaid claims for 17 states.
NHSN = National Healthcare Safety Network, the Centers for Disease Control and Prevention (CDC).
NDNQI = National Database of Nursing Quality Indicators
NDNQI/CALNOC = NDNQI, supplemented by data on the same measures for about 100 additional hospitals from CALNOC.
NVSS = National Vital Statistics System.
HENs = Outcomes data submitted in aggregate form by Hospital Engagement Networks.

These harm reductions are estimated to have saved at least $8.67 billion. Two very different sets of estimates of the cost savings associated with the reductions in adverse events described above are available from two different sources. Each set of estimates was derived using different methods and data sources and included different adverse events and PfP focus areas. One set was developed by the Evaluation Contractor and the other by AHRQ (Chapter 6). Due to data availability issues, neither of the estimates covers all the PfP focus areas and neither covers the full PfP time period for all the harm areas. Future evaluation reports will include updated estimates based on the entire PfP campaign period of performance.[10]

The Evaluation Contractor has estimated that harm reductions nationally have resulted in at least $8.67 billion in cost savings between 2010 (the year before PfP began) through either 2013 or mid-to-late 2014 (varying by measure). In this conservative estimate, many harms that were known to occur were not included in the cost savings estimate because there were no cost-per-event data from the literature that were well-matched to the type of harm being counted, so that the exclusions included ADEs, stage 1 pressure ulcers (more severe pressure ulcers were counted), falls with injury other than fractures, and harms that did not fall into the PfP focus areas. The estimate from AHRQ (which is less conservative) suggests savings may total $11.98 billion through 2013. Chapter 6 describes the methods underlying the two estimates and presents them broken out by area of harm. It is unclear what proportion of savings in either set of estimates is attributable to PfP.

[10] Agency for Healthcare and Research Quality (AHRQ) Saving Lives and Saving Money: Hospital-Acquired Conditions Update Interim Data from National Efforts to Make Care Safer, 2010-2014. Rockville, MD; December 2015. Available at http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2014.html

The environment of concurrent activity toward harm reduction complicated attribution to any one initiative. The substantial improvements to patient safety have occurred at a time when a number of initiatives to address patient safety problems were underway. The PfP campaign recognized most of these initiatives as aligned with its work including Quality Improvement Organization (QIO) initiatives, payment reforms, state-level initiatives, and private-sector initiatives. Many of these initiatives were likely aided by the spread of information fostered by PfP and vice versa. PfP actively coordinated with and encouraged these other efforts as partners in harm reduction. This active work with partners, along with the fact that the work toward similar goals was simultaneous, makes it very difficult to determine the contribution of any particular initiative to the national decrease in harm.

Although the PfP campaign was designed to be a national effort that would work synergistically with all of these other harm-reduction efforts, policymakers are still interested in understanding what portion of the reduction in patient harms can be attributed to the campaign. To understand this, the Evaluation Contractor examined the effect of a major component of the PfP campaign—the effect of the HENs.

PfP successfully reached hospitals through HENs, and a majority of those aligned with a HEN took action intended to improve patient safety to some extent due to their participation. There was a high rate of hospital participation through aligning with (committing to work with) a HEN. A majority of the surveyed hospitals that aligned with a HEN considered themselves to be engaged in the work with the HEN; a large number of hospitals nationally reported that they made changes to improve patient safety and readmissions as a result of their PfP engagement, citing it as more influential than other factors. From the perspective of many of the surveyed hospitals, the HENs were an effective approach for facilitating the spread of best practices and other resources. However, the ability to use that information to effect change and reduce harms varied from hospital to hospital.

Evidence of the impact of the HEN component on outcomes and costs is inconclusive. The Evaluation Contractor conducted three types of analyses, one that examined national data series for structural breaks coinciding with the inception of the HEN contracts, another that analyzed changes in HEN-supplied data over time, and a third that estimated the direct impact of the HEN component by comparing HEN-aligned and non-HEN-aligned hospitals, using a number of different sources of data and models. Across the three analyses, there was some evidence that HEN-aligned hospitals experienced changes associated with reduced patient harms during the timeframe of the PfP campaign. However, the results do not support the conclusion that the PfP campaign provided an independent impact to reduce patient harms in the targeted hospitals. Table 4 summarizes the quantitative findings for each harm area and for the aggregate measures of harm.

The ITS analysis of national database measures detected significant structural breaks associated with accelerating reductions in readmissions for both Medicare and Medicaid claims-based measures. In contrast, none of the HAC data series exhibited structural breaks toward greater improvements in Q1 2012, the first full quarter of the HEN operations. Two harm areas, pressure ulcers and CAUTI, exhibited structural breaks toward worsening rates in the national database measures. For all of the harm areas, the ITS analyses did not detect structural breaks coinciding with the beginning of the HEN contracts.

The ITS analysis of HEN-supplied data examined whether there was an overall reduction in harm and an increase in the rate of that reduction after the start of the HEN component. Few outcome measures showed this pattern, with readmissions and OB-EED showing a moderate proportion of measures meeting these criteria. However, this analysis was limited by the fact that the HEN-supplied data varied across HENs and across harms in the measures used, as well as in the duration, periodicity, completeness, and consistency of hospital reporting.

In one of the difference-in-differences analyses (Bayesian), which examined four measures from Medicare claims data, there was a moderate probability of a slight or small impact for two harms (VTE and catheter-related bloodstream infection [CRBSI], respectively), a moderate probability of substantial impact for one harm (pressure ulcers), and no impact for readmissions.[11] However, for all sets of the Bayesian estimates, the uncertainty interval around the impact estimates includes zero, so it is possible that the observed improvements are due to chance variation rather than HEN alignment. Frequentist difference-in-differences analyses of OB-EEDs and birth outcomes and of an aggregate measure of harms also found no measurable positive HEN impact. Analysis of Medicare FFS expenditures during and after a hospitalization found no difference in the change in those costs over time for beneficiaries admitted to HEN-aligned versus non-HEN-aligned hospitals. These results cannot be interpreted as evidence that the HEN component of the campaign did not work for any of the aligned hospitals, rather that little evidence was found that the HENs improved key outcomes for the average aligned hospital relative to changes being observed in non-HEN-aligned hospitals during this time period. Ironically, the more that PfP and its partners were able to reach all hospitals in the U.S., the less likely this type of analysis is able to estimate the impact of the HEN component on HEN-aligned hospitals relative to non-HEN-aligned hospitals.

Table 4—Summary of Quantitative HEN Results by PfP Harm Area

PfP Harm Area	ITS Analysis of National Trend (Change in *Trend–Q1 2012 Breakpoint*)	Difference-in-Differences Analysis (HEN-Aligned versus Non-HEN-Aligned)		ITS Analysis–Percentage of HEN Outcomes Measures Showing Both Overall Improvement and Accelerated Improvement Trend Following Start of HEN Component
		Type of Model	Result of Analysis of HEN Impact–Differential Change in HEN-Aligned versus Non-HEN-Aligned Hospitals (Data Source)[c]	
ADE	—	—	—	**17%**
VTE	**No Change**	**Bayesian**	**Moderate Probability of HEN Impact of 2-5%** Medicare	**8%**
Pressure Ulcers	**No Change** Medicare, Medicaid, NDNQI **Worsening Trend** NDNQI	**Bayesian**	**Moderate Probability of HEN Impact of 25%** Medicare	**9%**
Readmissions	**Improved Trend** Medicare, Medicaid	**Bayesian**	**No Evidence for HEN Impact** Medicare	**21%**
CLABSI	**No Change**	**Bayesian**	**Moderate Probability of HEN Impact of 5-10%** Medicare	**12%**

[11] "Moderate probability" is 60 to 80 percent.

Table 4—Summary of Quantitative HEN Results by PfP Harm Area				
PfP Harm Area	**ITS Analysis of National Trend (Change in *Trend–Q1 2012 Breakpoint*)**	**Difference-in-Differences Analysis (HEN-Aligned versus Non-HEN-Aligned)**		**ITS Analysis–Percentage of HEN Outcomes Measures Showing Both Overall Improvement and Accelerated Improvement Trend Following Start of HEN Component**
		Type of Model	**Result of Analysis of HEN Impact–Differential Change in HEN-Aligned versus Non-HEN-Aligned Hospitals (Data Source)[c]**	
CAUTI	**No Change** NHSN, Medicaid (2012[b]) **Worsening** NDNQI, Medicaid (2012[b])	—	—	**9%**
VAP	**No Change**	—	—	**17%**
SSI	**No Change** NHSN	—	—	**6%**
OB-EED	—	**Frequentist**	**No Evidence for HEN Impact** NVSS	**27%**
OB-Other	**No Change**	**Frequentist**	**No Evidence for HEN Impact** NVSS	**8%**
Falls	**No Change**	—	—	**11%**
Aggregate Measures of Harm	—	**Frequentist**	**No Evidence for HEN Impact on Total Medicare Expenditures** Medicare **No Evidence for Large (20% or more) HEN Impact on Aggregate Measure of Harm** MPSMS	—

Source: Evaluation Contractor's analysis of data sources as listed.

Note: Medicare claims data severely undercounts the number of harms relative to chart review, for the Bayesian analyses of VTE, pressure ulcers, and CLABSI (see Appendix D). However, the Evaluation Contractor is unaware of any evidence that the undercount results in a differential shortcoming for the HEN-aligned versus the non-HEN-aligned hospitals, and the readmissions analysis, OB-EED, OB-Other, and total Medicare expenditures analysis does not have this limitation.

[a]In the Medicaid claims data, two measures were tested: the CAUTI HAC measure and a measure for Hospital-acquired UTIs (HAUTI)—which includes many more infections some of which are not related to catheters. For HAUTI, the pediatric rate decreased while the adult rate increased through 2012. For the CAUTI HAC, the reverse was true.

[b]In the Medicaid claims data, ITS analysis of change in trend showed no change for HAUTI-pediatrics, and worsening for HAUTI-adults.

[c]High probability is 85 percent or more; moderate is 60 to 85 percent; and low is less than 60 percent probability.

Data source abbreviations (See Appendix B for a description of each source):

Medicare = Medicare FFS claims.

MPSMS = Medicare Patient Safety Monitoring System (chart review data abstracted by AHRQ).

NVSS = National Vital Statistics System.

HENs = Outcomes data submitted in aggregate form by Hospital Engagement Networks.

Since harms were reduced nationally during this time period, and over 80 percent of admissions nationally occurred in HEN-aligned hospitals, clearly improved safety within the HEN-aligned hospitals was a major part of the decrease. Results of the difference-in-differences analyses suggest that safety improved at an equal rate in the non-aligned hospitals. In addition, the ITS failed to show any relationship between HEN activity and outcomes. There are at least two potential conclusions one could draw from these analyses.

Potential Conclusion 1: *It is possible that the design and implementation of PfP, working with and alongside other stakeholders and payment incentives to improve patient safety and reduce readmissions across all hospitals, was successful in reducing harms across HEN-aligned and non-HEN-aligned hospitals alike.*

The analytic results do not show, but are consistent with, the possibility that PfP worked exactly as intended. That is, the vast majority of hospitals in the country—both aligned and not aligned with a HEN—received the support and encouragement they needed from PfP, with and alongside the other aligned stakeholders and payment incentives, to address their patient safety opportunities.[12] It is also possible that by influencing the actions of such a large number of HEN-aligned hospitals, that hospitals that were not aligned took similar actions, without any support or encouragement from PfP, so as not to fall behind in the marketplace.[13] If this were the case, analyses aimed at detecting a difference in harm reduction between HEN and non-HEN-aligned hospitals would underestimate the impact of the campaign. At the same time however, simply looking at trends in national rates since 2010 to identify the impact of PfP would miss the impact of other, unrelated, factors that influenced these outcomes. In fact, for many of the measures examined, the data series examined did not contain sufficient historical data to determine if there were significant structural breaks following the beginning of the PfP campaign. For the measures where such data existed, the results are mixed.

Potential Conclusion 2: *HEN-aligned hospitals received and acted on more information than non-HEN-aligned hospitals, but without measureable payoff in terms of reduced harm outcomes and readmissions during the PfP period.*

Under PfP, a great deal of information was shared with hospitals. This information fell into three categories:

- Evidence-based.
- Self-described experience of a hospital or HEN that appeared to be having success based on a decreasing trend of harm post-intervention.
- Anecdotal, based on networking with other hospitals, often facilitated by the HEN.

[12] Shafer et al. (2008) provide an example of another large-scale health quality improvement program where the focus of intensive assistance was on a relatively smaller set of hospitals, but the positive outcomes appeared to spread to other hospitals more broadly, possibly due to the broader roles of key partner organizations.

[13] Burns and Pauly (2002) conclude that "the history of the adoption of managerial innovations and new corporate forms in the hospital industry reveals the strong presence of local imitation and industrywide bandwagon effects." (pp. 134-135)

There is no record of the extent to which either the efforts of HEN-aligned hospitals or other hospitals' efforts to improve safety were based on these three categories of information. It remains possible that a great deal of work based on anecdotal content and content that was short of evidence-based practice was both needed and accomplished equally by HEN-aligned and non-HEN-aligned hospitals to achieve the national reductions in harm. It is also possible that HEN-aligned hospitals worked more than others on changes where the basis was short of evidence-based, and that this did not earn them any better results than non-HEN-aligned hospitals who may have invested less effort and resources on changes to reduce harms; or that hospitals were just less effective in implementing additional changes successfully; thus explaining the lack of relationship between HEN activities and change in outcomes.[14] Additionally, there is evidence of substantial variation in the performance levels exhibited by the HENs participating in the PfP campaign. Chapter 5 presents HEN-level analyses demonstrating the variability of HEN performance across specific metrics. Thus, the similarities demonstrated between HEN-aligned and non-HEN-aligned hospitals on average, may mask the possibility that the quality of implementation, and hence outcome performance, differed from HEN to HEN.

Considerations for Patient Safety Programs

While there has been marked improvement in overall patient safety in the last few years, the rate of inpatient harm was estimated at 12 harms per 100 discharges in 2013 (AHRQ National Scorecard) and readmissions are still common particularly among Medicare beneficiaries, about 18 percent of whom are readmitted after discharge. CMS can continue to build on the PfP experience in designing future support for quality and patient safety improvement.

Chapter 8 of this report provides many specific considerations for shaping future programs based on participant feedback and Evaluation Contractor's observation. These are considerations rather than recommendations because corroborating evidence that they would improve outcomes is unavailable. CMS may wish to consider incorporating into future programs:

- Features of PfP that HENs found helpful to their progress, such as PfP's bold aims and interim assessments, alignment with payment incentives and public reporting, emphasis on partnerships, and the assistance of national-level support contractors.
- Support for activities that HENs undertook that were associated with increased hospital engagement, such as peer-to-peer networking, skills training, and virtual consultation or coaching. These specific activities were associated with a greater likelihood of a hospital implementing operational changes targeting harm reduction compared to those that did not participate in these activities.
- Increased health information technology (HIT) support, since over one-fourth of survey respondents listed HIT support as a remaining need that would assist them with further harm reduction.
- Work to enhance the business case for harm and readmissions reduction, since resource limitations were cited by participants as a barrier to achieving greater levels of patient safety.

[14] Kaissi and Begun (2008) suggest that while imitation of exemplars (i.e., hospitals responding to the type of information described in the second bullet point above) can sometimes be efficient and beneficial, "managers should carefully take account of local conditions that may distinguish their own organizations from the exemplars." Although in many instances, HENs worked with hospitals to adapt interventions to their local environment, it is not clear how much overall success this activity produced.

1. Introduction

This is the interim evaluation report for the Partnership for Patients (PfP) campaign, a model designed by the Center for Medicare & Medicaid Innovation (CMMI) to harness a disparate group of providers, programs, and initiatives to greatly reduce patient harms across the nation.

Overview of PfP

The PfP model was a test of whether it is possible to achieve national spread of proven practices to reduce hospital inpatient harms and readmissions across all United States (U.S.) acute care hospitals. The specific goals of PfP were to reduce preventable inpatient harms by 40 percent and readmissions by 20 percent by the end of 2014. These ambitious goals were chosen to propel aggressive action by the hospitals. The Centers for Medicare & Medicaid Services (CMS) designed the campaign to align new and existing programs across the U.S. Department of Health and Human Services (HHS), and align programs among other federal, state, local, and private partners in order to focus national attention on making care safer in U.S. hospitals. Specifically, PfP was designed to reduce readmissions, plus inpatient harms in nine adverse event areas (AEAs): adverse drug events (ADE), falls, pressure ulcers, catheter-associated urinary tract infections (CAUTI), central line-associated blood stream infections (CLABSI), surgical site infections (SSI), ventilator-associated events (VAE), venous thromboembolism (VTE), and other obstetric (OB) adverse events (OB-Other).[1-1] This selection process, which took into account both the current existence of evidence for effective improvement strategies, along with the presence of potential for further improvement, made PfP a test of spread, rather than a test of a specific set of interventions.[1-2] Given the burden of harm on patients and families, the intent was primarily to spread pre-existing knowledge of how to reduce harm and to shorten the long period of time typically required to diffuse new knowledge.

This final program evaluation assesses the extent to which these goals were met. It also gauges whether the campaign met the standard of budget neutrality needed to justify continuation of this innovative model of care in the context of the Affordable Care Act (ACA). This requires a showing that CMMI's investment in the PfP campaign model generated improvements in the quality of care without increasing spending, reduced spending without compromising quality, or both.[1-3]

PfP was launched to the general public and to hospitals across the nation in April 2011, with the release of a webinar series by the National Quality Forum (NQF) that focused on prevention of inpatient harms. In parallel, hospitals and healthcare professionals were encouraged to sign a pledge to reduce preventable patient harms. PfP implementation intensified during 2012, after CMS awarded contracts to Hospital Engagement Networks (HENs) and support contractors in late 2011.

The campaign consisted of three engines for change: CMMI itself; its federal partners; and state, local, and private partners. CMMI contracted with HENs and led other national initiatives, including the Community-based Care Transitions Program (CCTP) and the Strong Start for Mothers and Newborns Initiative (Strong Start). Its federal partners included existing and newly developed programs within CMS and other federal

1-1 A tenth harm area was added beginning in 2012, when obstetrical early elective deliveries (OB-EEDs) were removed from the larger category of OB harm, and tracked independently.

1-2 The literature sources HHS used for the selection process and the estimation of preventable harms and potential lives and costs saved are listed at the end of this AHRQ document found online (undated), available at http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/pfp/interimhacrate2013.pdf

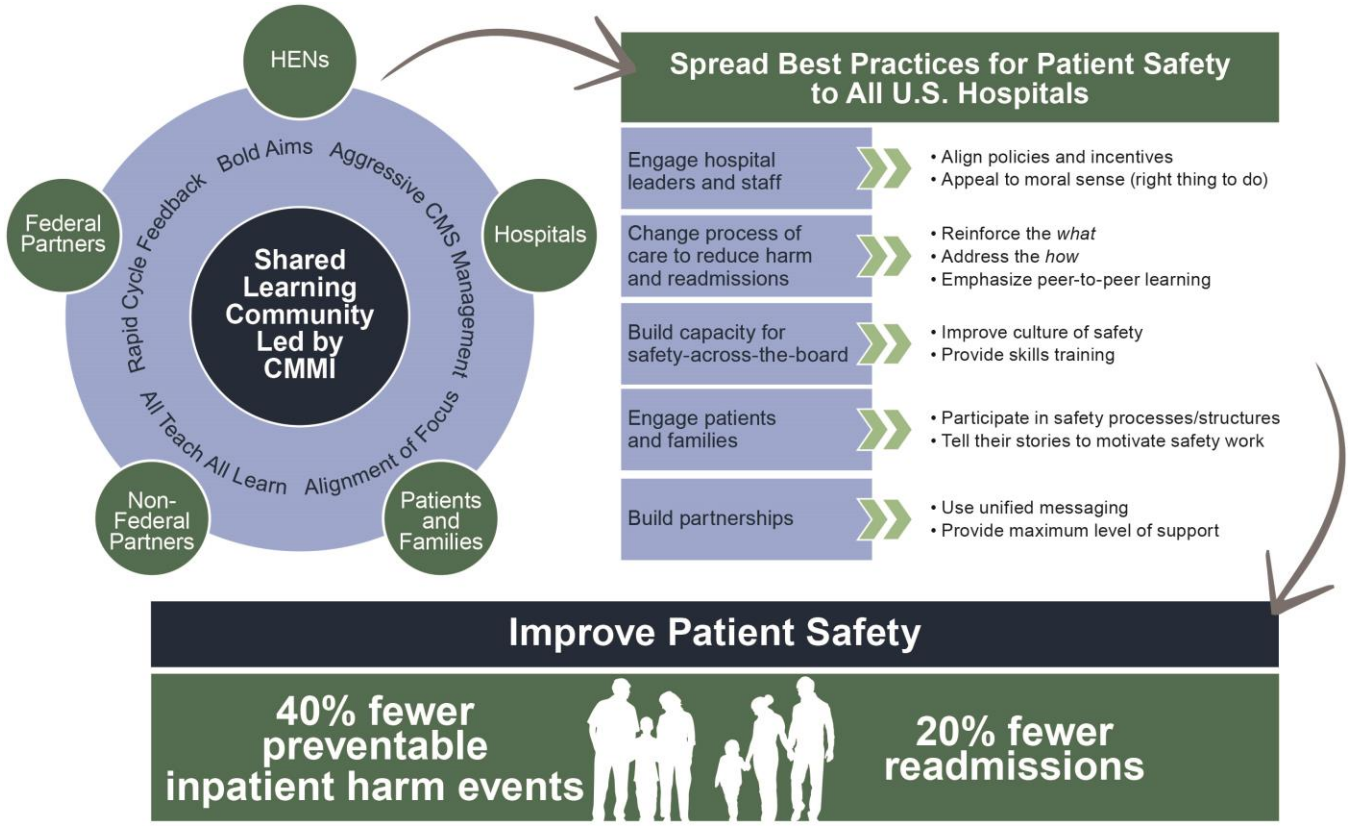
1-3 Affordable Care Act, Public Law 111-138, March 23, 2010, 124 STAT Section 393.

agencies, including the Quality Improvement Organization (QIO) program, the Administration on Aging, the Centers for Disease Control and Prevention's (CDC's) National Healthcare Safety Network (NHSN), the Agency for Healthcare Research and Quality (AHRQ), the Health Resources and Services Administration (HRSA), and the U.S. Office of Personnel Management (OPM). State, local, and private partners included a complex network of state departments of health, professional associations, unions, long term care facilities, healthcare providers, researchers, and employers.

How PfP was Designed to Achieve its Goals

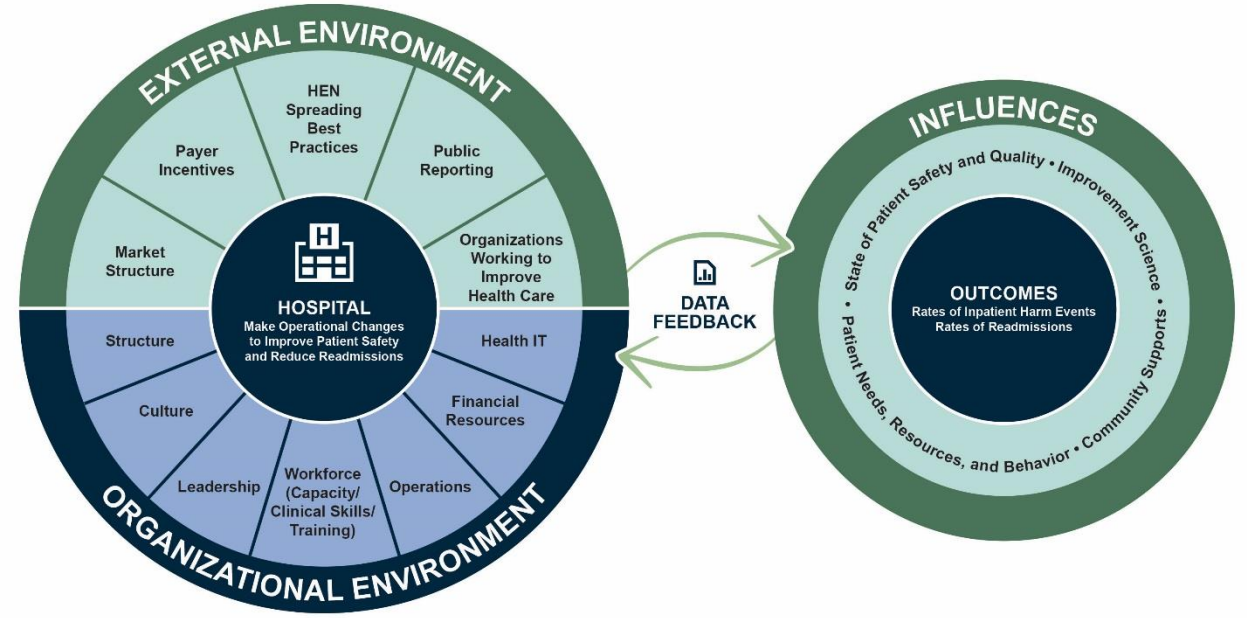
PfP was designed consistent with best practices for implementation of large-scale quality campaigns, and made use of principles for influencing change in complex adaptive systems, as illustrated in Figure 1-1.

Figure 1-1—The PfP Campaign Strategy



Given that hospitals are the key focal points for change under PfP, Figure 1-2 provides a model for change at the individual hospital level. The rest of the report will evaluate various aspects of change from both the campaign level (Figure 1-1) and the hospital level (Figure 1-2). The HEN is listed as only one component of the hospital's environment in Figure 1-2; however, the HENs worked to leverage and partner with the other components of the environment as much as feasible, and worked not only to influence the hospital's operational changes directly, but also to influence the data feedback loop, and to influence changes through its organizational environment.

Figure 1-2—Logic Model of Change at the Hospital Level



The Shared Learning Community

A shared learning community, led by CMMI and facilitated by several support contractors, included the CMS-funded HENs; federal partners (such as AHRQ and CDC); non-federal partners (e.g., March of Dimes and American Nurses Association [ANA]); patients and families (through a network nurtured by a Patient and Family Engagement Contractor [PFEC] funded by CMS); and hospitals that participated primarily through aligning with HENs as well as attending many national-scale virtual learning events under PfP. Please see Appendix E for a description of the improvement contributions from the federal and non-federal partner organizations.

The key tenets of the learning community were:

1. ***All teach, all learn***: Most learning events were designed to feature hospitals and HENs sharing their experiences and lessons learned for the benefit of other hospitals and HENs. Other events featured national-level experts in the PfP focus areas. PfP leadership also participated in learning as well as teaching, working with support contractors to adjust the learning event processes based on participant feedback intended to improve the usefulness of the events.
2. ***Alignment of focus***: PfP worked to achieve alignment of focus toward its goals with other drivers and contributors to the spread of patient safety improvement, including payment programs; reporting programs; key networks (such as rural hospitals supported by HRSA); incentives by other payers (such as the OPM Federal Employees Health Benefits [FEHB] Program); and work led by non-federal partners (such as the American Congress of Obstetricians and Gynecologists [ACOG] and its work to reduce maternal harm).
3. ***Rapid-cycle feedback***: PfP shared monthly formative feedback reports prepared by the Evaluation Contractor with and among the HENs in order to keep them apprised of their improvements, as well as their performance relative to other HENs. HENs excelling in a particular focus area were recruited to share insights with others. PfP leadership also conducted several full rounds of "one-on-one" calls with the 26 HENs, discussing the HENs' progress towards goals and any issues they were

encountering, and providing real-time management feedback. HEN rapid-cycle feedback to hospitals was also a key component of the campaign's strategy to spread best practices.

The Spread of Best Practices for Patient Safety

To spread best practices, PfP aimed to first engage hospitals nationally in the effort to improve patient safety and readmissions, and then both to disseminate best practices targeted to the focus areas and to build hospitals' capacity to achieve safety across the board. Building and strengthening partnerships was considered a key approach to providing maximum support for improvement at the large scale required to achieve goals, for harmonizing messaging around patient safety, and for promoting sustainability for a future that might not include HENs.

HENs and HEN-Aligned Hospitals

As noted above, the HENs were the major technical assistance arm of PfP and were directly funded by the campaign. HENs included large and small organizations of different types, all of which had existing relationships with hospitals on which to build. The largest HEN was the AHA/HRET HEN, which was composed of over 1,700 hospitals and worked closely with 31 state hospital associations (SHAs) to achieve the PfP campaign goals. Other HENs developed from hospital systems (e.g., Ascension Health [Ascension]), SHAs (e.g., New Jersey Hospital Association/Health Research and Educational Trust of New Jersey [New Jersey]), and other organizations connected with hospitals in multiple states (e.g., VHA, Inc. [VHA], Premier Healthcare Alliance [Premier], and the University Health System Consortium [UHC].)[1-4]

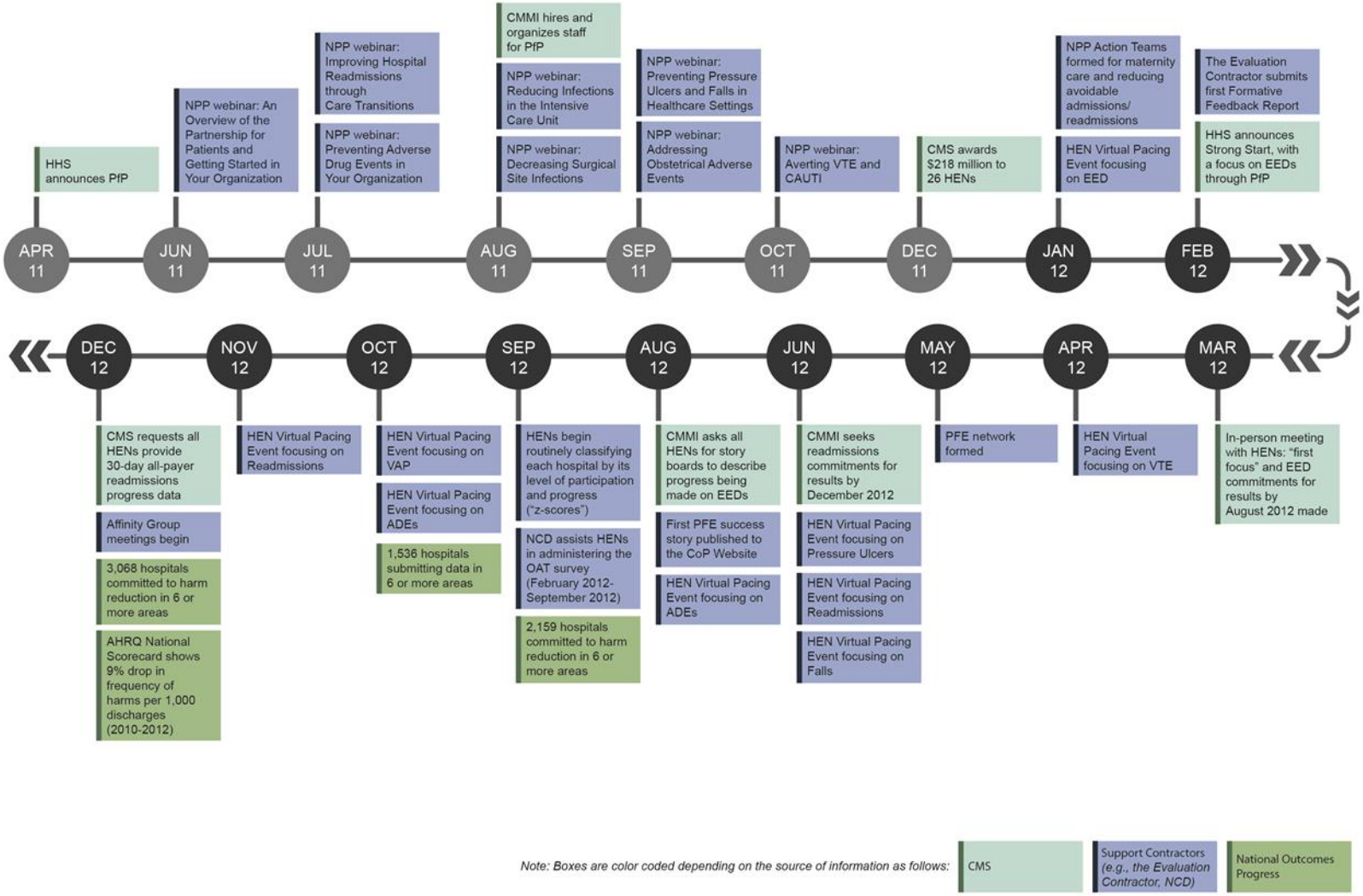
HENs were tasked with recruiting hospitals to participate in the PfP campaign by committing to reducing patient harms, participating with the HEN and other contractors, and actively working toward achieving the PfP goals. Hospitals that officially committed to working with a HEN in the PfP campaign are referred to as HEN-aligned. HEN-aligned received collaboration, support, and feedback from the HEN they were aligned with (see chapter 3 for further details). The HENs were encouraged to recruit aggressively, and ultimately over 70 percent of U.S. short-stay acute care hospitals (over 3,700 hospitals) agreed to participate with them in the campaign. However, the influence of the PfP campaign spread beyond these hospitals. Learning opportunities were open to everyone interested, regardless of whether the hospital was HEN-aligned or not. While non-HEN-aligned hospitals may have also worked with HENs and other contractors in an unofficial capacity, the extent to which this happened is unknown. However, the extent of support and feedback received from the HENs was likely to be less than those received by HEN-aligned hospitals.

Timeline of PfP Campaign

PfP sponsored an intensive and complex set of interventions. Figure 1-3 and Figure 1-4 provide timelines for 2011–2012 and 2013–2014, respectively, to describe the campaign's implementation and evolution.

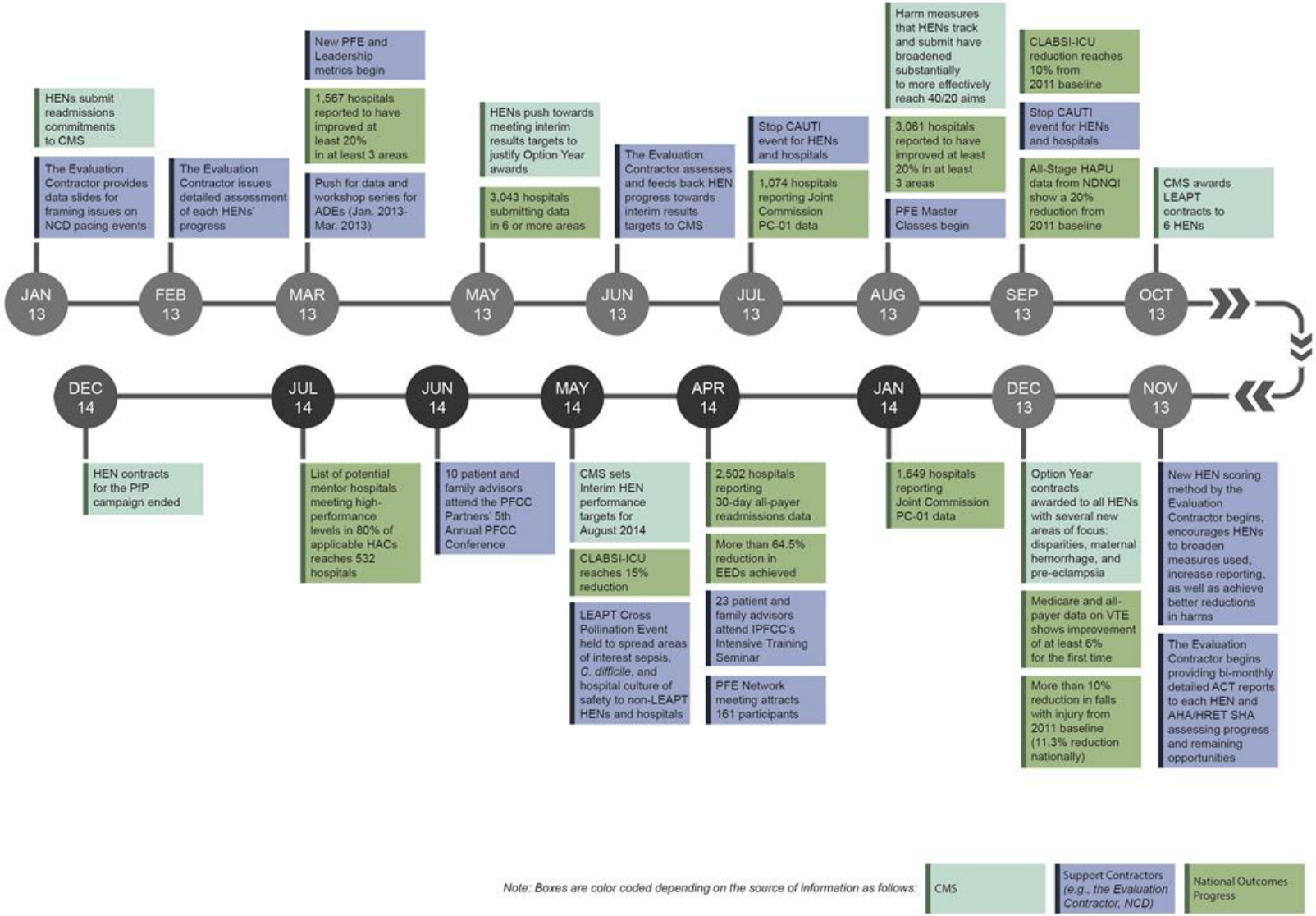
1-4 A full list of HEN names and abbreviations, as well as the distribution of HEN hospitals within states, can be found in Appendix A.

Figure 1-3—Timeline of PfP Events, 2011-2012



Note: NPP=National Priorities Partnership, CoP=Community of Practice, and OAT=Organizational Assessment Tool

Figure 1-4—Timeline of PfP Events, 2013-2014



Note: TJC PC-01=The Joint Commission Prenatal Care-Early Elective Delivery, HAPU=Hospital-Acquired Pressure Ulcer, LEAPT=Leading Edge Advanced Practice Topic, PFCC=Patient & Family Centered Care, IPFCC=Institute for Patient and Family-Centered Care,

Introduction to the Evaluation

This is the interim evaluation report for the PfP campaign, which uses a diverse set of mixed methodologies and equally diverse set of data sources. Table 1-1 shows characteristics of PfP (some of which are unusual in intervention/evaluation cycles) that collectively challenged the evaluation's ability to attribute change to the campaign. The national scale of PfP, coupled with the partnerships with other stakeholders targeting the same areas, are the biggest factors that make it challenging to assess the progress that would have been made in its absence. The Evaluation Contractor employed multiple data sets and methodologies—mixing qualitative and quantitative analyses—in an effort to understand more fully the effects of PfP on harm reduction. The implications of harm reduction for healthcare costs are also addressed, providing input to U.S. Department of Health and Human Services (HHS) in its assessment of whether the investment in PfP paid off for taxpayers.

Table 1-1—PfP Characteristics That Are Atypical and Challenged the Evaluation

Characteristic	PfP Characteristics Atypical for Intervention/Evaluation Cycles	Implications for the Evaluation
Scale	All patients within acute care hospital inpatient settings nationally	Along with the other characteristics below, limits the ability to construct a comparison group that is known to be free of the intervention's effects
Partnership	Major strategic component	Partner organizations typically serve wider constituencies beyond HEN-aligned hospitals – helps PfP spread to *all* hospitals but impedes evaluation's ability to know who was affected by PfP through the partnerships
Concurrent Work with Other Initiatives Targeting Similar Objectives	Widespread and actively promoted for many of the focus areas	Difficult to disentangle effects of PfP from effects of concurrent initiatives
Definition of Intervention	Locally defined	Complicates understanding of what was learned from PfP
Change to Participation During Funded Period	Dynamic–continual entry and exit	Creates multiple choices for how to identify the "treated" HEN-aligned group
Measurement Philosophy	Allow local choice in measures and flexibility in baselines and reporting periods to encourage widespread participation	Undermines usefulness of HEN-supplied data for the evaluation

Evaluation Overview

Research Questions

The three primary research questions CMS put forth for the evaluation are (chapters containing results are in parentheses):

- To what extent did inpatient harm and readmissions decrease nationally during 2011–2014, consistent with the PfP goal, and what costs were associated with this decrease? (Chapters 2 and 6)
- To what extent did PfP play a role in the decrease? (Chapters 3, 4, 6, and 7)

Additional questions addressed in the report include:

- To what extent was PfP successful in generating hospital engagement and action designed to prevent harm? (Chapter 3)
- Are there unintended consequences of PfP? (Chapter 3)
- Did some types of hospitals, or hospitals in some types of HENs, benefit more from PfP than others? (Chapter 5)
- What can stakeholder feedback on PfP operations, along with analysis of results, suggest for future program planning? (Chapters 3 and 8)
- What does the evaluation experience suggest for future evaluations of large-scale improvement programs? (Chapter 8)

To address these questions, the evaluation included descriptive and impact analyses, described below.

Specific Components of the Evaluation

The national scale of the PfP campaign and its expansive goals called for a broad array of analytic strategies. The Evaluation Contractor performed the following analyses, which are presented with the major methods and data sources used in these analyses:

- ***Assessed changes in national database measures*** on NHSN data, National Database of Nursing Quality Indicators® (NDNQI®) data, and CMS Policy Group data, using statistical process control (SPC) charts, and interrupted time series (ITS) analysis for significant changes in rates and trends in order to monitor progress in specific AEAs.[1-5] (Chapter 2)
- ***Compared PfP trends to those for QIOs*** using QIO data for Arizona, California, Florida, and Ohio and Medicare claims-based outcome measures to determine whether there were identifiable differences in trends across hospitals choosing to participate in the PfP relative to those working with a QIO, or choosing to do both or neither. (Chapter 2)
- ***Assessed the relationship between changes in hospital behaviors and patient outcomes*** with descriptive analysis of differences in mean change in rates of harms and readmissions between hospitals that newly implemented several patient safety-related processes compared with other hospitals that did not implement such processes. (Chapter 3)
- ***Identified unintended consequences*** using qualitative analyses to identify the extent to which the PfP campaign may have produced positive or negative consequences not intended by the program. This analysis utilized interview data from the Evaluation Contractor's hospital site visits and interview data collected by the Evaluation Contractor during fall 2014 HEN interviews. (Chapter 3)

[1-5] NDNQI® is a registered trademark of the American Nurses Association (ANA). NDNQI® data were supplied by ANA. The ANA disclaims responsibility for any analyses, interpretations, or conclusions.

- ***Assessed differences in HEN-aligned vs. non-HEN-aligned hospital group trends in national measures*** using propensity score reweighting with difference-in-differences regression to determine whether or not the HEN-aligned hospitals exhibited greater reductions in patient harms relative to a similar group of non-HEN-aligned hospitals using Medicare claims data. (Chapter 4)
- ***Performed Bayesian analysis of Medicare claims-based measures*** to improve estimates of the certainty around impacts, and reduce reliance on more strict distributional assumptions in frequentist statistical analyses. (Chapter 4)
- ***Assessed the changes in HEN-level outcome and process measures*** using ITS analysis and repeated measures analysis to evaluate the extent of change reported in the HEN-level data. (Chapter 4)
- ***Evaluated the impact of the Strong Start Initiative*** using difference-in-differences regression and standard econometric analyses with control variables to determine the impact of Strong Start on rates of obstetrical early elective deliveries (OB-EEDs), cesarean section (C-section) deliveries, and neonatal intensive care unit (NICU) stays. Counties with a high proportion of HEN-aligned births were contrasted against counties with low proportions of HEN-aligned births using OB-EED data from the National Vital Statistics System (NVSS), developed by the Medicaid Medical Directors Network (MMDN). (Chapter 4)
- ***Assessed differences in dose-response analysis across each harm area*** using ITS, repeated measures analysis stratified by different levels of engagement and activity, and survival model techniques. Using HEN-level data, the analysis evaluated whether or not HENs that exhibited greater levels of activity by their constituent hospitals also experienced greater reductions in harm. (Chapter 5)
- ***Assessed changes in HEN-level measures*** by preparing SPC charts for HEN-level outcome measure data to determine if these measures exhibited evidence of special cause variation. (Chapter 5)
- ***Analyzed spillover effects*** using descriptive statistics and regression analysis on Medicare claims-based outcome data, to compare outcomes across non-HEN-aligned hospitals that indicated contamination and non-HEN-aligned hospitals that did not indicate contamination in order to develop an estimate of bias generated by contamination. Data on spillover were drawn from the Evaluation Contractor's Survey on Prevention of Adverse Events and Readmissions. (Chapter 4)
- ***Evaluated the effect of hospital engagement in HEN initiatives*** using propensity score reweighting and difference-in-differences regression to provide a form of dose-response analysis by examining relationships between hospital level of engagement and outcomes, where engagement is measured from self-reported survey data. This analysis utilized Medicare claims data, AHA survey data, and the Evaluation Contractor's 2012 and 2014 Survey on the Prevention of Adverse Events and Readmissions. (Chapter 5)
- ***Estimated changes in costs incurred during 90-day and 180-day follow-up periods*** using difference-in-differences regression to estimate the cost differential for patients admitted to HEN-aligned vs. non-HEN-aligned hospitals utilizing Medicare claims data. (Chapter 6)

Overview of Methods

To evaluate the PfP program, the Evaluation Contractor used multiple data analysis methods. These include: examining trends over time using SPC charts, descriptive analysis of how the campaign was implemented using qualitative and survey data, comparing changes over time in hospital implementation, comparing changes in outcomes across HENs using ITS, and comparing the changes in outcomes between hospitals that aligned with HENs with those that did not align with HENs. An overview of the major methods that were used in the evaluation is described as follows (please refer to Appendix C and D for more detail on the methodologies employed in this evaluation):

SPC Charts

Quality improvement data are inherently time-delimited, and as such, analyses of the timing and direction of changes are particularly relevant to understanding the success of any quality improvement campaign. Importantly, virtually all quality metrics in healthcare will exhibit some degree of variability over time due to random changes in patient mix, the hospital environment, and other factors external to the facility. At other times, healthcare quality measures will change because a substantively important process has been altered in a nonrandom fashion (e.g., a quality improvement intervention was implemented to reduce patient harms). SPC charts provide one method for assessing time-series data and identifying the time periods during which the changes in a series might reasonably be considered random variation versus special cause variation.

The SPC methodology differentiates special cause (i.e., non-random) variation from random variation by examining the observed variation in the data series for a patient harm, and taking into account the number of underlying observations that are included in the harm rate at any given time point. Based on these factors, the SPC chart is defined with control limits that represent probabilistic boundaries for random variations around the average, or "true" rate. Where the sample size at a given time point is larger, the control limits are defined so that smaller deviations from the average level of the measure will signal special cause variation. Over time, as the measure rate varies, each data point can be assessed for its evidence of random versus special cause variation. Consecutive points are also assessed for evidence of trends in the measures, or for evidence that the average level of a measure has changed. When such evidence is observed, the control limits are adjusted to reflect the change in processes that produced a new rate.

Descriptive Analyses

To understand how the campaign was implemented, the Evaluation Contractor collected qualitative and survey data to describe how the campaign was implemented. The analysis seeks to explain how the PfP campaign strategy modeled in Figure 1-1 was implemented, leading to an understanding of how the learning community operated and how the HENs perceived the campaign.

Qualitative Analysis and Coding

Qualitative data were collected through several sets of interviews, all of which used semi-structured interview protocols and were documented in full detail and systematically coded for analysis. Relevant frameworks for studying dissemination and implementation were reviewed and used to shape data collection protocol and analyses (see Appendix C). For the largest set of interviews, the notes were coded and analyzed with the aid of Atlas.ti software version 7.5.4. Coders were trained and cross-checked to achieve consistent coding, and all coded notes also were completely reviewed by one of two senior staff members, who also spot-checked each other to ensure coding was consistent. Emerging themes were shared, discussed, and challenged for robustness. In report text, the number of respondents associated with a particular theme is cited in parentheses for transparency as to the extent of support for the statement.

ITS Analysis

The Evaluation Contractor used a second form of longitudinal analyses, ITS, to assess the improvements in specific measures reported by PfP HENs and national database measures. ITS analysis is a robust quasi-experimental design for assessing the impact of an intervention on a time-series outcome, and is typically used to assess the statistical significance of changes in an outcome at a predefined `time` point. The ITS framework examines the level and trend in the measure rate before and after the beginning of the intervention and determines whether the measure exhibits a statistically significant change, called a structural break, after the intervention begins. Structural breaks can take the form of a change in average level, a change in the

trend over time for the outcome, or both. The detection of a structural break following the implementation of a quality improvement intervention would be evidence of an intervention effect. In the case of the PfP campaign, the expected impact of the interventions is to reduce the average levels of patient harms and/or change the slope of the trend toward greater reductions in patient harms.

Importantly, the ITS framework makes several key assumptions related to the processes influencing the outcome of interest. First, the ITS model assumes that absent the intervention, the trend in the outcome measure would have continued along its pre-intervention path. Thus, where quality measures are already declining prior to the intervention in question, the impact of the intervention must produce a significantly greater decline or downward shift in average level in order to be detected. Second, without additional time-varying explanatory variables included in the model, the ITS framework assumes that all other factors relevant to the outcome are not changing. To the extent that unmeasured explanatory factors are changing, the ITS model may detect a structural break due to factors other than the intervention of interest or may fail to detect a structural break due to one intervention as a result of the confounding by additional interventions. As a corollary to the second assumption, the third assumption of the ITS framework is that when multiple interventions are implemented, each can be distinctly identified in time from the other interventions. Where multiple interventions overlap or are implemented at the same time, the ability of ITS analysis to identify the portion of a structural break attributable to each intervention will be weakened.

Survival Analysis—Cox Proportional Hazard Model

As a portion of the dose-response analysis, the Evaluation Contractor examined the length of time required for measures of patient safety to improve and achieve the goal for reduction in patient harms. This type of data is also known as time-to-event data, and can be used to determine the rate at which an event occurs over time. For example, did PfP outcomes tend to reach goal reductions quickly at the beginning of the campaign and less quickly later? This might imply that measures where gains could be made easily were targeted early on, while more difficult patient harms took longer to reduce. In contrast, did the rate of measures achieving the goal increase over time? This pattern might imply a learning effect wherein the HEN-aligned hospitals became better at reducing patient harms over time.

A survival analysis, also known as a Cox proportional hazard regression model, was used to assess the relationship between the percentage of hospitals participating in different types of HEN activities and the number of months required for the measures to reach the goal reductions of the PfP campaign. The Cox regression provides an estimate of the change in the likelihood of a measure experiencing the event of interest, or achieving goal in this case. Therefore, where higher percentages of hospitals participating in HEN activities are associated with reduced time to meeting the goal, this would constitute evidence of a dose-response relationship among the PfP HENs.

Impact Analyses

Impact analyses are designed specifically to determine the extent to which outcomes can be attributed to an intervention or program. In this evaluation, the Evaluation Contractor cannot measure the impacts of the overall campaign, because of the broad-based work with partner organizations to influence all U.S. hospitals. Instead, the design focused on approaches to measure the impacts of the HEN component of the campaign. The following types of impact analyses were used for the evaluation:

Difference-in-Differences Regression Analyses

An impact analysis that yields an estimate of what would have happened in the absence of the campaign, also called a counterfactual, requires a comparison group. The Evaluation Contractor has not been able to establish a counterfactual for the federal agency and non-federal partner components of PfP, as they affect all hospitals nationwide. The impact analysis focuses on the HEN component of PfP, for which an approach to establishing a counterfactual exists. Thus, in the remainder of this report, discussions of the impacts of PfP refer specifically to the impacts or effects of the HEN component of PfP. Ironically, to the extent that PfP's goal of influencing *all* U.S. hospitals is met, the evaluation has less chance of identifying an impact from the HEN component.

In order to compare the amount of improvement that occurred in HEN-aligned hospitals to the amount of improvement that might have been expected had those hospitals not worked with a HEN, the Evaluation Contractor constructed a comparison group of non-HEN-aligned hospitals and compared the change in the outcomes between the two groups. When appropriate, the comparison group was created using propensity score reweighting for each regression analysis that used Medicare claims data. The following provides a brief summary of the propensity score reweighting techniques used with the difference-in-differences regression analyses (a more detailed explanation of the construction of the comparison group is in Appendix D):

Propensity Score Reweighting

Propensity score reweighting produces a comparison group of non-HEN-aligned hospitals that are similar to HEN-aligned hospitals on observable characteristics of hospitals and their patients, by assigning different weights to non-HEN-aligned hospitals depending on their similarity to HEN-aligned hospitals. More weight is given to non-HEN-aligned hospitals that are more similar to HEN-aligned hospitals, and less weight is given to non-HEN-aligned hospitals that are less similar to HEN-aligned hospitals.

The propensity score reweighting approach used in this evaluation consists of two steps. First, a propensity score model is estimated, in which participation in PfP is a function of relevant hospital characteristics hypothesized to affect HEN participation and the PfP outcomes. Second, weights are constructed from the estimated propensity scores to weight the non-HEN-aligned (comparison group) hospitals in order to make the hospitals similar to treatment (HEN-aligned) hospitals on observable characteristics.

Separate propensity models were estimated for each AEA, for readmissions, and for total Medicare cost. Baseline levels of each AEA (2011 was the baseline year) and readmission rates (baseline year was 2010) were used.[1-6] A different sample of hospitals was used for each analysis as well, with the number of hospitals (or patients, depending on the analysis) provided with the results of each analysis. Hospital data using Medicare claims were the most complete, nationally, while hospital data based on survey results or the Medicare Patient Safety Monitoring System (MPSMS) data had smaller sample sizes.

Difference-in-Differences Comparison Group Analyses

The Evaluation Contractor compared change over time among HEN-aligned and comparison group hospitals using a regression-based difference-in-differences approach. This approach removes biases in estimated impacts that could result from any time-invariant differences between the treatment and comparison groups that remain after propensity score reweighting, or from any factors unrelated to the HENs' work with hospitals that affect changes in patient safety and readmissions for both groups (such as other CMS quality improvement efforts underway at the same time as PfP).

[1-6] A sensitivity test was conducted for total Medicare costs, considered the most sensitive indicator, to see if using 2010 versus 2011 as the baseline changes the results, and it did not change the results.

Difference-in-Differences Regression Specification After Adjusting for Non-Aligned Hospitals Influenced by PfP Using Medicare Data

For each AEA, a difference-in-differences model was fit relating individual patient outcomes (e.g., CAUTI, SSI) to PfP and a set of demographic and other control variables. The Evaluation Contractor estimated difference-in-differences models as follows:

$$y_{iht} = \delta_{y[t]}\, PFP_h + \gamma\, t_t + \phi\, w_i + \theta\, x_i + \beta\, z_h + \varepsilon_i, \quad (1)$$

where the outcome variable, y_{iht}, is measured for a hospital discharge (i) occurring in hospital h in quarter t. The variable PFP_h is a dummy variable for whether the hospital where the discharge occurred was HEN-aligned and the coefficient $\delta_{y[t]}$ estimates the effect in year y of hospital alignment with a HEN, t_t is a vector of quarterly dummy variables indicating the quarter in which the observation took place, and the estimated coefficients ($\gamma = [\gamma_1,\gamma_2,\ldots,\gamma_T]$) control for secular trends in the outcome variable. The regression model also includes patient-level covariates that control for demographics, patient risk factors, and characteristics of the hospital where the discharge occurred. The regression model also includes hospital-level characteristics as a vector of hospital dummies (z_h)—also known as hospital fixed effects—to control for all hospital-specific observed and unobserved factors that are stable over time. Finally, ε_i is an error term. Equation 1 follows the approach used for the difference-in-differences comparison group analyses, described in detail in Appendix D. In addition to examining results separately for 2012 and 2013, results for 2012 and 2013 combined were also examined.

Equation 1 is specified for the discharge-level outcomes (CRBSI, pressure ulcers, and VTE), but the Evaluation Contractor used a hospital-level model for the readmissions outcome. In this hospital-level model, there was one observation for each hospital for each year and the outcome variable ranged from zero to one. The patient demographics and comorbidities were aggregated to the hospital level for each year, but otherwise the analyses were similar to those for the discharge-level outcomes.

Difference-in-Differences Regression in a Bayesian Model Using Medicare Data

A detailed description of the Bayesian model is provided in Appendix D. In general, the Bayesian difference-in-differences analysis is analogous to the model described above, except for the following changes:

- Assigning a "prior distribution" to each model parameter translated the model into the Bayesian framework. A prior distribution describes one's understanding of a parameter before any data are taken into account. The noninformative priors used in this analysis for the overall impact of the HEN component across all HEN-aligned hospitals assign equal prior probability to all possible values of a parameter, equivalent to making no prior assumptions at all.
- Reframing provided fuller information on the likelihood that the campaign improved outcomes of different magnitudes.
- To make Bayesian computation feasible, the Evaluation Contractor fit the model using a two-stage formulation and then fit the first stage risk adjustment model (a discharge-level, propensity-score-weighted linear regression). The risk-adjusted, hospital-quarter-level output from stage 1 is used as data in stage 2, which estimates the effect of the HEN-aligned intervention in a Bayesian difference-in-differences framework.[1-7]

[1-7] The model adjusted for race, sex, and relevant comorbidities.

Difference-in-Differences Analyses Using Vital Records

Vital Records are available only at the county level rather than for individual hospitals; therefore, a slightly different analytic approach from previous difference-in-differences analyses was used. The Evaluation Contractor estimated the number of births in each county that occurred in hospitals aligned with HENs and created two main analytic groups: counties where 90 percent or more of births occurred in HEN-aligned hospitals (the "treatment" group), and counties where 50 percent or less of births occurred in HEN-aligned hospitals (the "comparison" group). The Evaluation Contractor then compared the two types of counties using a similar difference-in-differences approach, except that for this analysis it was not possible to use propensity score reweighting or matching methods to make the comparison counties "match" the treatment group counties. There were too few counties in the potential comparison group (in most counties, most births occur in HEN-aligned hospitals), and those counties differed from the HEN-aligned counties in several important ways (Kranker et al. 2014).

Difference-in-Differences Analyses of Expenditures, Using Medicare Data

The difference-in-differences regression analysis used for analyzing Medicare expenditures is similar to the model described above in equation (1). For each type of expenditure, a difference-in-differences model was fit relating individual patient outcomes (e.g., total expenditures, index admission expenditures, post-discharge expenditures) to PfP HEN alignment and a set of demographic and other control variables. This is a discharge-level analysis, where observations from HEN-aligned hospitals received a weight of one and observations from non-HEN-aligned hospitals received the propensity score-based weight for each outcome. Medicare claims data for calendar years 2009 through Q1 2014 were used to conduct this analysis.

Difference-in-Differences Analyses Using MPSMS Data

MPSMS is a project sponsored by AHRQ in which specific adverse events are identified in a national sample of hospitalized patients through abstraction of their hospital medical charts (despite the name, the data include patients covered by all insurers, not only Medicare). The inpatient medical records are those that were sampled as part of CMS' Inpatient Quality Reporting (IQR) Program for calendar years 2009 through 2013. After merging with Medicare claims data (for patients covered by Medicare) and the AHA survey data, the MPSMS data were used in a difference-in-differences regression analysis to compare chart-abstracted adverse event rates in HEN-aligned and non-HEN-aligned hospitals. Because of the smaller number of hospitals included in the MPSMS data, propensity score weighting could not be used. Differences between the HEN-aligned and non-HEN-aligned hospitals were adjusted for by using regression models that included patient- and hospital-level characteristics.

Survey Analyses

The Evaluation Contractor used descriptive analysis of data from the Hospital Survey on Prevention of Adverse Events and Reduction of Readmissions (targeting a nationally representative sample of hospitals) for three types of analyses: results at a single point in time, differences between HEN-aligned and non-aligned hospitals, and national-level change over time.

Simple *t*-tests were used to compare statistical significance of differences (1) between 2012 and 2014 rounds of the survey, and (2) between HEN and non-HEN respondents. In addition, a set of regressions were used to explore the relationship between hospitals' self-reported level of engagement with PfP (ranging from full engagement to none) by PfP focus area, and outcomes in the Medicare claims data. For each of four outcomes (central venous catheter-related blood stream infection [CRBSI] rate, stage 3 and higher pressure ulcers, post-operative VTE rate, and readmissions), the analysis identified whether there was a difference in

change over time in the outcome between the group at each engagement level (fully, moderately, minimally) versus those not engaged at all. Hospitals not aligned with a HEN were grouped with HEN-aligned hospitals that reported they were not engaged at all, and propensity score reweighting was used in making comparisons.

Descriptive analysis of data from the Survey of Hospitals' Participation in Patient Safety Activities (target was all HEN-aligned hospitals) was used to describe the level of engagement with different types of HEN activities, and the extent to which hospitals responded due to PfP and other factors, by harm area. Simple logistic regression analysis (regression analysis that does *not* control for other factors) was used to examine the data for a potential association between HEN characteristics and hospitals' level of participation and response.

Major Data Sources

The following section provides detail on the major data sources used for this report and notes key strengths and weaknesses as they pertain to the measures used in these analyses.

Outcomes Data

- ***AHRQ National Scorecard data***—Used to provide an overview of national progress on harm reduction, these data on the incidence of adverse events are predominately drawn from a nationwide sample of inpatient charts selected as part of CMS' Hospital Inpatient Quality Reporting (IQR) program (MPSMS data, described next).[1-8] Statistical testing has not been performed by AHRQ to establish whether changes over time are significant and is not available to the Evaluation Contractor at this time.
- ***MPSMS data***—Used in analysis of composite rates of inpatient harm, these data contain 21 adverse event measures for a sample of 50 to 70 medical charts per hospital, for 800 hospitals per year (with a different sample every year) and for each year from 2009 to 2013. Although the sample includes 17,000 to 34,000 charts per year, it is not large enough to detect moderate differential HEN impacts on individual harm indicators using difference-in-differences analyses of individual harm indicators, so composite measures were used. The sample is restricted to patients with four conditions: acute myocardial infarction (AMI), heart failure (HF), pneumonia (PN), and surgical patients (as defined by the Surgical Care Improvement Project [SCIP]).
- ***Medicare claims data***—Used in analysis of readmissions and AHRQ-defined Patient Safety Indicators (PSIs), these data are available for the entire Medicare fee-for-service (FFS) population. Measures based on claims data typically undercount the actual number of harm events. Appendix B, highlights this point by showing a direct comparison of harm identification within the Medicare FFS claims data analyzed for the report against harm identification from the MPSMS data, for Medicare patients present in both databases, finding much lower levels of harm identification in the claims data.

[1-8] For more about the method underlying the scorecard, see http://www.ahrq.gov/professionals/quality-patient-safety/pfp/#methods. The outcomes in the scorecard that *are* not from the medical chart-based MPSMS are OB harms (claims data on two Patient Safety Indicators [PSIs] from the AHRQ Healthcare Cost and Utilization Project are used) SSIs (estimated by the CDC), and several components with the "Other hospital-acquired conditions" category, which use HCUP data for several additional PSIs.

- ***Vital Records data***—Used for analysis of OB-EEDs and selected birth outcomes; the county-level data used in the analysis contain most births in the U.S. and an extensive set of perinatal measures. The records contain some clinical detail but less than the ideal set of details for isolating all justified early deliveries.
- ***HEN-level data***—Data self-reported by HENs in the aggregate for their aligned hospitals were used for the ITS analysis. Additionally, a small set of non-HEN-source measures was also included in the ITS analysis. Typically, these are all-payer measures and closely represent the HENs' focus for improvement; however, the data the HENs submitted vary as to the specific measures provided, the time periods provided, and the hospital reporting levels contained in the aggregates (which differ over time among some HENs).
- Other national-level data used in analysis of national trends only (Chapter 2):
  - **NDNQI and Collaborative Alliance for Nursing Outcomes (CALNOC) data**: The NDNQI and CALNOC data contain rates of falls, pressure ulcers, CAUTI, CLABSI, and ventilator-associated pneumonia (VAP). Falls and pressure ulcer measures from NDNQI and CALNOC are more sensitive and accurate than claims-based measures for these areas due to their use of a surveillance method. However, NDNQI and CALNOC are voluntary databases in which hospitals pay to participate. Therefore, it is likely that hospitals that placed a higher value on participation and improvement and may have dedicated more resources to this issue are disproportionately represented in the database. Only aggregated data were available to the evaluation contractor, not data for individual hospitals.
  - **NHSN Data**: Data on healthcare-associated infections (HAIs) are provided by hospitals to the CDC. The CDC provided to PfP, in aggregate form, measures of CAUTI and CLABSI in ICU units, and SSIs. These data are nationally representative of hospitals in Medicare's Inpatient Prospective Payment System (IPPS), since reporting is required, and the measures are risk-adjusted. However, because the CDC only provides aggregated data to the Evaluation Contractor, it is not possible to use clustered standard errors as would be appropriate in statistical testing of these data; this limits the use of these data for the evaluation.
  - **Medicaid claims data**: Claims data from 17 states (Alabama [AL], Arkansas [AR], Connecticut [CT], Georgia [GA], Iowa [IA], Illinois [IL], Indiana [IN], Louisiana [LA], Maryland [MD], Michigan [MI], Missouri [MO], Montana [MT], Oregon [OR], Pennsylvania [PA], South Dakota [SD], Vermont [VT], and Wyoming [WY]), including FFS (all states) and managed care (GA, IN, and MI), were used to calculate AHRQ-defined PSIs (pressure ulcers, CRBSI, VTE, OB trauma, and injury to neonate), AHRQ-defined Pediatric Quality Indicators (PDIs), CRBSI, and hospital-acquired conditions (HACs) (CAUTI/hospital-acquired urinary tract infection [HAUTI] for adults and pediatric populations). These data have a longer lag than other sources and only extend through 2012 or, for some indicators, Q1 2013. Since present on admission (POA) data are not available for these data, the current analyses assumed all diagnoses were *not* POA.

Survey Data

- ***Hospital Survey on Prevention of Adverse Events and Reduction of Readmissions***—A nationally representative sample of short-term acute care hospitals was surveyed in 2012 (response rate 70 percent), and respondents were resurveyed in 2014 (response rate 67 percent), to learn about their experience with PfP, how their patient safety processes and infrastructure had changed, and what factors influenced them during this period. A total of 1,136 hospitals (47 percent of those originally sampled) responded to both rounds of the survey and were included in analyses of changes in practice.

- ***Survey of Participation in Patient Safety Activities***—Almost all HEN-aligned hospitals had the opportunity to complete this survey and the response rate among those surveyed was 72 percent, providing insight on a large percentage of HEN-aligned hospitals' participation and response to PfP, albeit on a limited set of questions.

Other Data

- ***HEN activity timelines***—In fall 2014, all HENs completed a spreadsheet listing their specific major initiatives and partnerships during PfP, along with the timing, and some characteristics of partnerships.
- ***HEN interview data***—HEN representatives were interviewed in summer 2014 and fall 2014, each time using a different semistructured interview protocol, with different foci. The topics that were covered and were of most relevance to this report were their experience with the PfP learning community (summer 2014) and their intervention strategies for each focus area (fall 2014).
- ***Non-federal partner interview data***—The Evaluation Contractor conducted 30-minute interviews with representatives of 19 organizations identified as non-federal partners at the national level, using a semi-structured protocol, to understand the nature of the alignment, coordination, and collaboration with PfP that took place (winter 2015).
- ***Review and selective email follow-up of federal partners' written summaries of their role in PfP***—These written descriptions were submitted in advance of a federal partners retreat held in November 2014. Email contacts were used to clarify some of the descriptions.

Limitations of Analyses and Caveats

Descriptive and Qualitative Analyses

Descriptive qualitative analyses presented in this report are helpful to understanding the patterns of implementation, outcomes, and cost reduction, but do not imply causal relationships.

The interviews that underlie many of the qualitative analyses in this report are limited to the information and ideas that the respondents chose to convey. For the *Survey on Prevention of Adverse Events and Reduction of Readmissions* there was a single respondent per hospital; although the instructions requested they consult others if they did not know the answer, there is no way to know if they did that, such that there may be error or bias due to their specific role and sphere of understanding of their hospital. There is also no way to know if their recollections were faulty, leaving out or misremembering potentially significant pieces of information. Similarly, individuals by nature bring their own biases and perspectives to the table; observations about factors that helped and hindered their progress were likely colored by their perspective to some degree. The qualitative research used research-based frameworks wherever possible, but no single framework fit PfP's multilevel implementation and dissemination focus exactly, requiring judgments to be made about which factors from multiple frameworks were most applicable.

The Evaluation Contractor's estimate of costs associated with national reduction in harms and readmissions is a very rough, conservative estimate based on the Evaluation Contractor's identification of the best available estimates of per-event cost from the literature or new analyses, which nevertheless have varying methodologies, strengths, and weaknesses. The many studies of hospital-acquired conditions all agree that harm is costly, so to recognize this, an estimate of savings from the national decrease in harms was prepared despite the lack of consistency in available cost-per-event estimates and the lack of ideal data or sources from which to calculate costs that match well with the available data for events averted (method and rationale described in Chapter 6). Many of the per-event estimates are based on hospital data alone, and thus do not

include professional fees; they would be higher if these fees were included (an exception is the pressure ulcers estimate, which is broader-based). In most cases, hospital cost data rather than expenditure data was used, however, for the falls with injury estimate the figure is based on expenditure data only, which again is typically lower than cost. The estimate was unable to include a cost estimate associated with reduction in ADEs and HACs that were not within the PfP focus areas of harm, because of lack of a good enough match between the measures used in the available data to count averted events and the measures used in the available cost estimates. In addition to the limitations of the cost-per-event figures, the counts of events averted were very incomplete for falls and OB harms other than OB-EED, where data from only about 1/4 to 1/3 of the nation's acute care hospitals (depending on the measure) were available that matched the cost-per-event estimates.

Impact Analyses

There are three major challenges facing the impact analyses: (1) measurement of inpatient adverse events; (2) estimation of impacts (i.e., the changes in outcomes that can be attributed to PfP), given the voluntary, nationwide nature of the campaign; and (3) power to detect an impact.

Measurement: Adverse events may be detected through self-reports by clinicians or patients, real-time direct observation of patient care, review of medical records, or specific diagnostic codes in administrative claims data. Many of the analyses reported here—those that used Medicare data as a source—rely on administrative claims data, which often fail to detect adverse events found by medical chart review (Appendix E).[1-9]

Estimation of impacts: Many analyses reported here compare HEN-aligned hospitals to a comparison group of non-HEN-aligned hospitals. Based on survey responses, some of these non-HEN-aligned hospitals were affected by the national campaign. As noted above, the Evaluation Contractor has now analyzed the sample of hospitals that participated in a national survey in which they were asked whether PfP "as a national effort" influenced their patient safety actions. However, because of PfP's partnerships at the national and local levels, hospitals receiving support because of PfP through partner organizations may not know that PfP was involved and the accuracy of impact estimates based on differential rates of change between HEN-aligned and non-HEN-aligned hospitals may have weakened.

Power to detect an impact: The impact analyses presented in the report vary in their power to detect an impact. The data source best capable of identifying patient harms, since it uses chart review—the MPSMS data—also uses sampling methodology and despite including over 28,000 discharges each year, the power of analyses with this data set is seriously limited. For example, the analysis only has a 37.5 percent chance of detecting an impact of 5 percent in the composite measure of harm (the best case). Although Medicare data include nearly all FFS beneficiaries nationally, the rarity of most of the relevant harm events in the claims data (all except readmissions) continues to limit the power of the analysis to detect small changes.

[1-9] The difference in frequency of adverse events is even more apparent when the rate of adverse events from Medicare claims using AHRQ PSI measures is compared to the rate identified in the medical chart-based MPSMS, although the difference is due to a combination of different measure definitions as well as potential undercoding in the claims.

SPC Charts, ITS, and Dose Response Analyses

These same three major challenges affecting the impact analyses also affect the analyses of HEN-level data.

Measurement: HENs self-reported data on aggregate outcomes for participating hospitals, and on activities and partnerships they developed throughout the course of the campaign. Among the aggregate measures of PfP outcomes and processes, data were not contributed by all participating hospitals within a HEN and HENs did not identify which hospitals were included in each measure. HENs also did not report on a standardized set of measures, instead reporting on measures chosen internally by each HEN and its constituent hospitals. The measures were not reported on a standardized schedule, resulting in some measures with monthly periodicity and others with quarterly periodicity. Additionally, the HENs and hospitals exhibited a high degree of variability in the volume and detail included when describing interventions and partnerships, resulting in a limited ability to identify reliable measures of PfP dosage or intensity.

As a result of these measurement aspects of the data, the changes in outcomes over time will likely be reflective of a combination of quality improvements in patient care, as well as the changing composition of the participating hospitals. Additionally, making direct comparisons across HENs can only be accomplished with a limited set of measures, called common measures. Measures reported on a quarterly basis may not reveal significant changes due to the reduced granularity of quarterly data relative to data reported on a monthly basis. Finally, attempts to assess the impact of HEN initiatives and partnerships will be attenuated to a greater degree among the adverse event areas where the data are less reliable as a result of the different ways HENs defined and reported on initiatives and partnerships.

Estimation of impacts: SPC charts do not compare the outcomes of a treatment group to those of a comparison group. Rather, the methodology is used to assess the timing and nature of variability in an outcome measure that may accurately be defined as nonrandom, or special cause, variation. While special cause variation may coincide with the implementation of specific intervention activities, without detailed knowledge of the context in which the HEN initiative was implemented, the SPC chart cannot provide attribution directly to the PfP campaign.

ITS analysis is a robust quasi-experimental design for assessing the impact of an intervention on a time-series outcome and is typically used to assess the likelihood of a significant change in an outcome at a predefined point. Due to the large number of concurrent activities reported by the PfP HENs in every reporting period, this approach was not feasible. Where significant changes in the time series data were detected using ITS analysis, the results could not attribute the impact to a specific activity. Additionally, due to the limited detail regarding the nature and timing of HEN intervention activities, the ability to link specific interventions to outcomes was attenuated. Also, it is important to note that the majority of measures reported by the HENs did not include time series data prior to the PfP campaign (i.e., baseline data typically consisted of several months of data aggregated into a single data point). The lack of trended baseline data prevented he Evaluation Contractor from assessing whether or not structural breaks may have occurred coincident with the beginning of the HEN contracts.

Dose-response models are used to assess the relationship between the exposure level of a subject to a treatment and the change in the outcome exhibited as a result of that exposure. For the PfP campaign, the dosage of HEN interventions and hospital participation were related to the direction and magnitude of the changes in patient harm outcome measures. However, the unreliable nature of the dosage measures resulted in a weakened ability to detect a dose-response relationship.

Power to detect an impact: The HEN-level data were reported with differing periodicity from one HEN to another and from one AEA to another. HENs were also encouraged to stop reporting a measure if it proved unreliable or invalid, and to begin reporting more reliable and valid measures as they became available throughout the campaign. With respect to the analysis of data for SPC charts and interrupted time series analyses, the number of observations in a data series is directly related to the ability to obtain reliable estimates of change. Where measures were reported with less frequency (i.e., quarterly data as opposed to monthly data), and when measures were reported for less than the complete duration of the PfP campaign, the analyses have reduced power to identify statistically significant changes in the outcomes.

Overview of Report Structure

This introductory chapter is followed by five evaluation results chapters (Chapters 2–6) that focus on national trends in inpatient harms, engagement of hospitals and their implementation of operational changes, and a quantitative analysis of the impact of HENs (a key component of PfP) on observed outcomes:

Chapter 2 presents the results of analyses—from independent national data sources and from HEN-provided data—of the overall trends and reductions in patient harms since the beginning of the PfP campaign. Additionally, this chapter summarizes the results of a series of SPC charts that analyzed trends in data submitted by the HENs. Finally, the chapter provides qualitative analyses of major parallel programs that may have contributed, with PfP, to a reduction in patient harm rates as well as a quantitative comparison of QIO and PfP trends.

Chapter 3 contains two major sections. The first describes the work of the HENs as a major component of PfP, as well as the factors that affected their ability to spread best practices and their perception of key features of PfP. The second provides qualitative analysis of hospitals' engagement with PfP, the extent to which they made operational changes designed to lead to harm reduction, and the role of PfP in their changes. The chapter also includes a section on unintended consequences, as reported by HENs.

Chapter 4 presents several attribution analyses that assess the contribution of the PfP campaign to the national reduction in patient harms. First, a summary table of findings by AEA across different types of analysis is provided. This is followed with results of a Bayesian difference-in-differences model examining the possible differential impact of the PfP campaign on patient harm rates for each AEA. Next is an assessment of whether removing spillover, as measured in an Evaluation Contractor survey, enhanced the ability to detect impacts using the survey linked with Medicare claims data for the survey sample. Separate difference-in-differences analyses are presented on OB-EED and OB outcomes, and on composite rates of inpatient harm, using data from the NVSS natality files, and MPSMS data, respectively. Finally, the ITS results are discussed by AEA, and a repeated measures analysis is evaluated.

Chapter 5 examines differences in results and trends among different subgroups of HENs and hospitals, assessing differences in HEN and hospital engagement, HEN and hospital characteristics, and HEN activities. In addition, several examples of notable reductions in patient harms made by different HENs.

Chapter 6 presents an estimation of costs averted associated with the national reduction in patient harms, and as a result of the HEN component of PfP. The analyses include estimating national cost averted by applying per event cost estimates to the estimated reduction in patient harm events. The analyses also includes difference-in-differences estimates comparing HEN-aligned and comparison hospitals, with propensity score weighting using varying sampling strategies and for two different time periods (expenditures within 90 days and 180 days from the Index Discharge Date). Total Medicare expenditures are

then analyzed in several different dimensions, including costs by facility type and provider type, in two different time periods (those within 90 days and 180 days from Index Discharge Date).

Finally, the report concludes with two chapters dedicated to synthesizing these results: **Chapter 7**, a discussion of results, and **Chapter 8**, which presents the Evaluation Contractor's considerations for future programs and for approaches to evaluation of large-scale quality improvement programs.

In addition to the chapters provided in the main body of this report, there are five accompanying appendices (A through E) that contain detailed descriptions of the analytical methods, data sources, and results. Appendices also provide details of analysis that were deemed interesting to some audiences, although not central to the purpose of this interim evaluation report.

2. National Trends in Inpatient Harms

The Partnership for Patients (PfP) campaign was designed to improve hospital safety across the United States (U.S.) hospital system. This chapter documents the national trends in hospital adverse events and readmissions before and after the start of the campaign, reflecting the progress that has been made in each of the PfP focus areas. It does not speak to the causes driving the changes.

The data are from multiple sources, with varying methods for measuring outcomes. As a result, the evidence for change in national trends may vary within adverse event area (AEA) by data source and method used. Table 2-1 provides an overview of which measures demonstrated improvement, or lack thereof, in rate across available all-payer, Medicaid, and Medicare data sources. In addition, the table illustrates how trends have changed over time by data source, and it presents the evidence for any statistically significant changes in levels or trends of the outcome measures.[2-1]

Rates of harm by most measures and across many data sources have noticeably reduced over the last 5 years. Harm area by harm area, available all-payer measures show reductions consistent with achieving the PfP goal by the end of 2014 for adverse drug events (ADE), venous thromboembolism (VTE), ventilator-associated pneumonia/ventilator-associated event (VAP/VAE), central line-associated blood stream infection (CLABSI), and obstetrical early elective delivery (OB-EED). Medicaid measures for catheter-associated urinary tract infection (CAUTI), all-payer measures for surgical site infection (SSI), and all-payer and Medicare measures for pressure ulcers show progress consistent with achieving the goal.[2-2] This is not meant to be a definitive assessment. Each measure has limitations as discussed in Chapter 1; data from the Agency for Healthcare Research and Quality's (AHRQ's) National Scorecard are final for 2013, and the Centers for Disease Control and Prevention's (CDC's) National Healthcare Safety Network (NHSN) data for CAUTI, CLABSI, and SSI are aggregated at the hospital level without adjustments for clustered standard errors. In addition, for CAUTI and SSI, the earliest NHSN data available were in 2012, when the Hospital Engagement Network (HEN) contracts were awarded. As a result, control limits were based on data from Q1 2012 forward, rather than any period prior to the HEN contracts.

Tests for changes in trends after the PfP campaign began often showed no acceleration of trend, with the notable exceptions of readmissions, where the trend accelerated downward after the inception of PfP according to the available measures. However, the data were insufficient to analyze all of the important trends in this manner (such as the component harms of the AHRQ National Scorecard), and there are limited pre-PfP period data points for some measures.

Table 2-1 summarizes the evidence of changes in all-payer, Medicaid, and Medicare measures during the course of the PfP campaign. The section following Table 2-1 provides the detailed analysis on a measure-by-measure basis. The chapter concludes by examining the evidence of concurrent PfP-aligned programs that were working to reduce patient harms.

2-1 Appendix E includes PfP focus area context tables, which provide detailed summaries of the strategies employed to reduce harms by PfP and its partners.

2-2 The PfP goal is 17.6 percent reduction in harms (40 percent reduction of preventable harms, with 44 percent assumed preventable based on AHRQ's review of the literature pre-PfP), and there are roughly 4 years between the start of PfP and the end of 2014.

Table 2-1—Summary of Evidence for Changes in Harms for All-Payer, Medicare, and Medicaid Populations, by Outcome Measures				
Measure	**Source**	**Evidence for Improvement (Cell entries specify the method showing improvement)**	**Evidence for No Change (Cell entries specify the method showing no change)**	**Evidence for Worsening (Cell entries specify the method showing worsening)**
OB-EED				
Early Elective Delivery Rate (The Joint Commission [TJC] Perinatal Care [PC]-01)	HENs (time period varies from 2010 through 2014)	*z*-tests of pre/post rates		
OB-EED (Non-Medically Indicated Early Term Singleton Birth)	Vital Statistics Q1 2009 to Q4 2013	U' chart Pre/post comparison of levels Pre/post comparison of trends		
Readmissions				
Medicare Fee-for-Service (FFS) 30-Day All-Cause Readmissions	Medicare Claims January 2009 to December 2014	U' chart Pre/post comparison of levels Pre/post comparison of trends		
Index Admissions Stays	Medicare Claims January 2009 to December 2014	*X* chart Average linear trend		
Medicaid 30-Day All-Cause Non-OB Readmissions, Adults	Medicaid Claims Q1 2009 to Q4 2012	U' chart Pre/post comparison of levels Pre/Post comparison of trends		
Medicaid 30-Day All-Cause Non-OB Readmissions, Children	Medicaid Claims Q1 2009 to Q1 2013		U' chart Pre/post comparison of levels Pre/post comparison of trends	
30-Day All-Cause All-Payer Readmissions	HENs (time period varies from 2010 through 2014)	*z*-tests of pre/post rates		
All-Payer, All-Cause, All Diagnosis 30-Day Readmissions per 100 Index Admissions	AHRQ CY 2010 to CY 2013	Observed reduction in rate		
ADE				
Combined ADEs (5) in MPSMS Adverse Events/HACs per 1,000 Discharges	AHRQ National Scorecard Q1 2010 to Q4 2013	U' chart Average linear trend		
ADEs per 1,000 Discharges	AHRQ National Scorecard CY 2010 to CY 2013	Observed reduction in rate		

Table 2-1—Summary of Evidence for Changes in Harms for All-Payer, Medicare, and Medicaid Populations, by Outcome Measures				
Measure	**Source**	**Evidence for Improvement (Cell entries specify the method showing improvement)**	**Evidence for No Change (Cell entries specify the method showing no change)**	**Evidence for Worsening (Cell entries specify the method showing worsening)**
VTE				
Perioperative Pulmonary Embolism (PE) or Deep Vein Thrombosis (DVT) per 1,000 Surgical Discharges (AHRQ Patient Safety Indicator [PSI]-12)	Medicare Claims Q2 2011 to Q4 2014	U' chart Average linear trend Pre/post comparison of levels	Pre/post comparison of trends	
	Medicaid Claims Q4 2009 to Q4 2012		U chart Pre/post comparison of levels Pre/post comparison of trends	
	HENs (All-Payer) (time period varies from 2010 through 2014)	*z*-tests of pre/post rates		
VTE per 1,000 Discharges	AHRQ National Scorecard CY 2010 to CY 2013	Observed reduction in rate		
VAP/VAE				
VAP per 1,000 Ventilator Days	National Database of Nursing Quality Indicators (NDNQI®)[2-3] Q1 2011 to Q4 2013	U' chart Average linear trend	Pre/post comparison of levels Pre/post comparison of trends	
VAP per 1,000 Discharges	AHRQ National Scorecard CY 2010 to CY 2013	Observed reduction in rate		
CAUTI				
CAUTI Standardized Infection Ratio (SIR) for Intensive Care Units (ICUs)[a]	NHSN Q1 2012 to Q4 2014		X chart Average linear trend Pre/post comparison of levels Pre/post comparison of trends	
CAUTI Device Utilization Ratio (ICUs)	NHSN	Average linear trend		
CAUTI Device Utilization Ratio (Non-ICUs)	NHSN	Observed reduction in rate		
CAUTI SIR (Non-ICUs), Adjusted for Device Utilization	NHSN	Observed reduction in rate		
CAUTI per 1,000 Catheter Days	NDNQI Q1 2011 to Q4 2014		U' chart Pre/post comparison of levels	Pre/post comparison of trends

2-3 NDNQI® is a registered trademark of the American Nurses Association (ANA). NDNQI® data were supplied by ANA. The ANA disclaims responsibility for any analyses, interpretations, or conclusions.

Table 2-1—Summary of Evidence for Changes in Harms for All-Payer, Medicare, and Medicaid Populations, by Outcome Measures				
Measure	**Source**	**Evidence for Improvement (Cell entries specify the method showing improvement)**	**Evidence for No Change (Cell entries specify the method showing no change)**	**Evidence for Worsening (Cell entries specify the method showing worsening)**
CAUTI (Hospital-Acquired Condition [HAC]) per 1,000 Adult Discharges	Medicaid Q1 2009 to Q4 2012	Average linear trend	U' chart Pre/post comparison of levels Pre/post comparison of trends	
Hospital-Acquired Urinary Tract Infection (HAUTI) per 1,000 Adult Discharges	Medicaid Q1 2009 to Q1 2013		U' chart Pre/post comparison of levels	Pre/post comparison of trends
HAUTI per 1,000 Pediatric Discharges	Medicaid Q1 2009 to Q1 2013	Average linear trend	U' chart Pre/post comparison of levels Pre/post comparison of trends	
CAUTI per 1,000 Discharges	AHRQ National Scorecard CY 2010 to CY 2013	Observed reduction in rate		
CLABSI				
CLABSI SIR for ICUs[a]	NHSN Q1 2011 to Q4 2014	X chart Average linear trend	Pre/post comparison of levels Pre/post comparison of trends	
CLABSI per 1,000 Central Line days	NDNQI Q1 2011 to Q4 2014	U' chart Average linear trend	Pre/post comparison of levels Pre/post comparison of trends	
Central Venous Catheter-Related Blood Stream Infection (CRBSI) per 1,000 Discharges (AHRQ PSI-07)	Medicare Claims Q2 2011 to Q4 2014	U' chart Average linear trend	Pre/post comparison of levels Pre/post comparison of trends	
CRBSI per 1,000 Adult Discharges (AHRQ PSI-07)	Medicaid Claims Q1 2009 to Q4 2012	U/U′ chart Pre/post comparison of levels	Average linear trend Pre/post comparison of trends	
CRBSI per 1,000 Pediatric Discharges (AHRQ Pediatric Quality Indicator [PDI]-12)	Medicaid Claims Q1 2009 to Q1 2013	U' chart Pre/post comparison of levels Pre/post comparison of trends	Average linear trend	
CLABSI per 1,000 Discharges	AHRQ National Scorecard CY 2010 to CY 2013	Observed reduction in rate		
Falls				
Falls with Injury per 1,000 Patient Days	NDNQI and the Collaborative Alliance for Nursing Outcomes (CALNOC) Q1 2011 to Q4 2014	U' chart Average linear trend	Pre/post comparison of levels Pre/post comparison of trends	

Table 2-1—Summary of Evidence for Changes in Harms for All-Payer, Medicare, and Medicaid Populations, by Outcome Measures				
Measure	**Source**	**Evidence for Improvement (Cell entries specify the method showing improvement)**	**Evidence for No Change (Cell entries specify the method showing no change)**	**Evidence for Worsening (Cell entries specify the method showing worsening)**
Falls per 1,000 Patient Days	NDNQI and CALNOC Q1 2011 to Q4 2014	U' chart Average linear trend	Pre/post comparison of levels Pre/post comparison of trends	
Post-Operative Hip Fracture per 1,000 Discharges (AHRQ PSI-08)	Medicare Claims Q2 2011 to Q3 2014		U chart Average linear trend Pre/post comparison of levels Pre/post comparison of trends	
Falls per 1,000 Discharges	AHRQ National Scorecard CY 2010 to CY 2013	Observed reduction in rate		
Other Obstetrical Harm (OB-Other)				
Injury to Neonate (AHRQ PSI-17)	HENs (time period varies from 2010 through 2014)		z-tests of pre/post rates	
	Medicaid Q1 2009 to Q1 2013	Average linear trend	U chart Pre/post comparison of levels Pre/post comparison of trends	
OB Trauma, Vaginal Delivery with Instrument (AHRQ PSI-18)	HENs (time period varies from 2010 through 2014)	z-tests of pre/post rates		
	Medicaid Q1 2009 to Q4 2012		U' chart Average linear trend Pre/post comparison of levels Pre/post comparison of trends	
OB Trauma, Vaginal Delivery without Instrument (AHRQ PSI-19)	HENs (time period varies from 2010 through 2014)	z-tests of pre/post rates		
	Medicaid Q1 2009 to Q4 2012	Pre/Post comparison of trend	U' chart	Pre/post comparison of levels
Admission to Neonatal Intensive Care Unit (NICU)	Vital Statistics Q1 2010 to Q4 2013			U' chart Average linear trend Pre/post comparison of levels Pre/post comparison of trends
APGAR Score < 6 at 5 Minutes	Vital Statistics Q1 2009 to Q4 2013		U' chart	Pre/post comparison of levels Pre/post comparison of trends

Table 2-1—Summary of Evidence for Changes in Harms for All-Payer, Medicare, and Medicaid Populations, by Outcome Measures				
Measure	**Source**	**Evidence for Improvement (Cell entries specify the method showing improvement)**	**Evidence for No Change (Cell entries specify the method showing no change)**	**Evidence for Worsening (Cell entries specify the method showing worsening)**
Assisted Ventilation Required	Vital Statistics Q1 2009 to Q4 2013	U' chart Average linear trend Pre/post comparison of levels Pre/post comparison of trends		
Low Birth Weight	Vital Statistics Q1 2009 to Q4 2013		U' chart Pre/post comparison of levels Pre/post comparison of trends	
AHRQ PSI-18 and PSI-19 Combined	AHRQ National Scorecard CY 2010 to CY 2013	Observed reduction in rate		
Pressure Ulcers				
Hospital-Acquired Pressure Ulcers (HAPU) per 1,000 Assessed Patients, All Stages	NDNQI and CALNOC Q1 2011 to Q4 2014	**U' chart**	Pre/post comparison of levels	Pre/post comparison of trends
HAPU per 1,000 Assessed Patients, Stages 2+	NDNQI and CALNOC Q1 2011 to Q4 2014	U' chart Average linear trend	Pre/post comparison of levels Pre/post comparison of trends	
HAPU per 1,000 Assessed Patients, Stages 3+	NDNQI and CALNOC Q1 2011 to Q4 2014	Average linear trend	U' chart Pre/post comparison of levels Pre/post comparison of trends	
Pressure Ulcers per 1,000 Discharges, Stages 3+ (AHRQ PSI-03)	Medicare Claims Q2 2011 to Q4 2014	Average linear trend	U' chart Pre/post comparison of levels Pre/post comparison of trends	
	Medicaid Claims Q1 2009 to Q4 2012		U' chart Pre/post comparison of levels Pre/post comparison of trends	Average linear trend
Pressure Ulcers	AHRQ National Scorecard CY 2010 to CY 2013	Observed reduction in rate		

Table 2-1—Summary of Evidence for Changes in Harms for All-Payer, Medicare, and Medicaid Populations, by Outcome Measures				
Measure	**Source**	**Evidence for Improvement (Cell entries specify the method showing improvement)**	**Evidence for No Change (Cell entries specify the method showing no change)**	**Evidence for Worsening (Cell entries specify the method showing worsening)**
SSI				
SSI Abdominal Hysterectomy SIR[a]	NHSN Q1 2012 toQ4 2014		X chart Average linear trend Pre/post comparison of levels Pre/post comparison of trends	
SSIs per 1,000 Discharges (17 procedure types)	AHRQ National Scorecard CY 2010 to CY 2013	Observed reduction in rate		
All Harm				
AHRQ National Scorecard HACs per 1,000 Discharges	AHRQ National Scorecard CY 2010 to CY 2013	U' chart Average linear trend		

Source: The Evaluation Contractor's analysis of all-payer, Medicare, and Medicaid data. Due to a lag time in receiving Medicaid data, these analyses only use Medicaid claims through Q1 2013.

Note: The table summarizes evidence of changes in rates of adverse events and readmissions. It does not speak to the causes driving the changes. Where different methods generate conflicting results regarding improvement versus worsening of rates, the test result that is more robust and consistent with the observed trend is presented in bold-faced font.

Evidence for Improvement: A meaningful decrease in the rate of harms, where "meaningful" means that the decrease is statistically significant, and/or meets statistical process control criteria for a special cause decrease, and/or is large in magnitude (the latter only in cases where only aggregated data are available and were not able to be tested for significance)

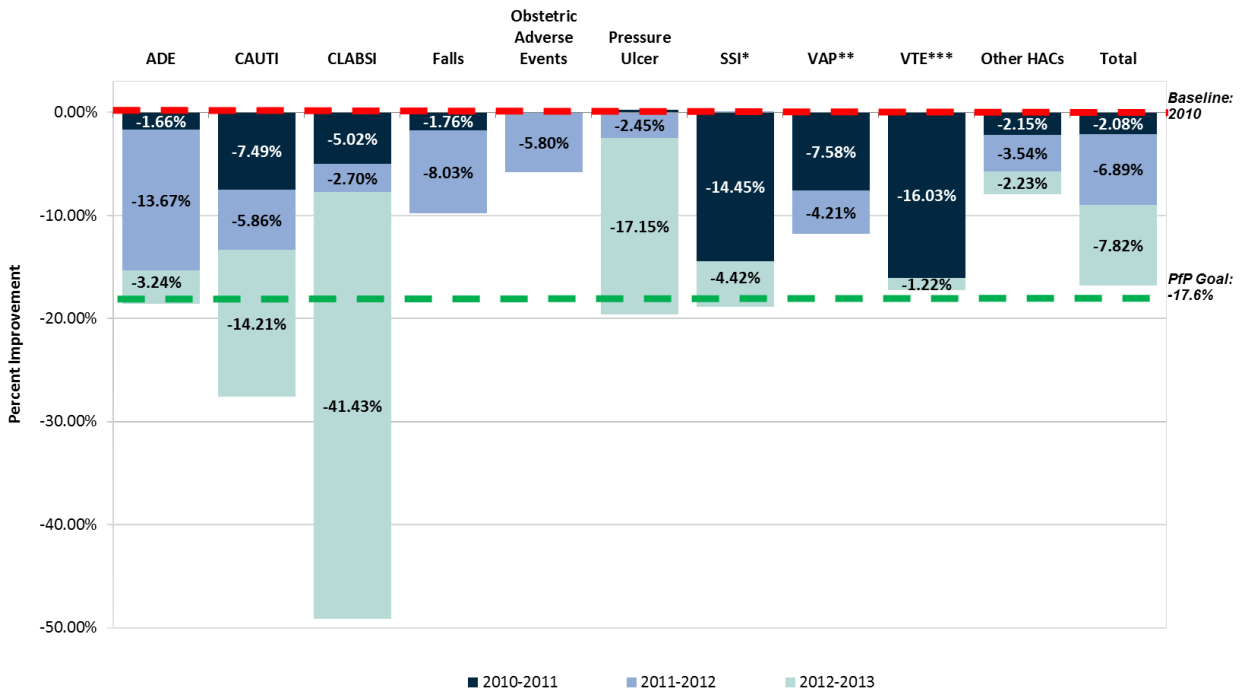
Evidence of No Change or Mixed Evidence: For these measures, clear evidence for improvement was either not present, or varied depending on which statistical test was used.

Evidence for Worsening: A meaningful increase in the rate of harms, where "meaningful" means that the increase is statistically significant, and/or meets statistical process control criteria for a special cause increase, and/or (in cases where only aggregated data are available) is large in magnitude.

[a]The CDC's NHSN data on SIRs were not provided to the Evaluation Contractor with clustered standard errors for use in statistical analyses. The ITS analyses presented in this chapter were performed under the assumption that discharges within hospitals are independent observations. If this assumption does not hold, then statistical tests for changes in trends will be liberal and results based on more appropriate clustered standard errors would provide more accurate and conservative results.

The AHRQ National Scorecard data, which provides the only available way to examine all the harms together, suggests that the nation is approaching the goal for overall harm rate reduction. Given that AHRQ used a denominator of inpatient discharges for all of its measures, the estimated rates of different harm events can be summed to provide an indication of trends in overall harms. The overall rate of harm estimated by AHRQ has dropped from 145 per 1,000 discharges in 2010 to 120 per 1,000 discharges in 2013, a 17.24 percent decrease. Figure 2-1 illustrates AHRQ's National Scorecard data.

Figure 2-1—AHRQ's National Scorecard Estimated by Type of Harm and in Total, in Relation to PfP Goal



Rates (per 1,000 Discharges)	ADE	CAUTI	CLABSI	Falls	Obstetric Adverse Events	Pressure Ulcer	SSI	VAP	VTE	Other HACs	Total
Baseline 2010	49.50	12.21	0.55	7.94	2.50	40.31	2.93	1.16	0.85	27.30	145.25
Current 2013	40.30	8.85	0.28	7.20	2.36	32.51	2.38	1.12	0.71	25.14	120.86
Percent Change (2010-2013)	**-18.57**	**-27.56**	**-49.15**	**-9.25**	**-5.63**	**-19.33**	**-18.81**	**-3.56**	**-17.25**	**-7.91**	**-16.80**

Source: PfP's "AHRQ National Scorecard" or National HAC Rate: Updated with final 2013 Data on the Medicare Patient Safety Monitoring System (MPSMS) HACs, provided by Noel Eldridge, AHRQ Center for Quality Improvement and Patient Safety (CQuIPS), on October 21, 2014 and November 20, 2014.

Note: AHRQ's data covers 9 of 10 original PfP focus areas: ADE, CAUTI, CLABSI, falls, OB adverse events, pressure ulcers, SSI, VAP, VTE, and "all other HACs," (readmissions and OB-EEDs are not included).

*There was an increase of 0.07 percent from 2011 to 2012 in SSI, after a -14.45 percent decrease from 2010 to 2011. The increase from 2011 to 2012 is not represented in the figure.

**There was an increase of 8.23 percent from 2012 to 2013 in VAP events after the 11.79 decrease from 2010 to 2012. The total decrease from 2010 to 2013 was 3.56 percent. The increase from 2012 to 2013 is not represented in the figure.

***There was an increase of 32.00 percent from 2011 to 2012 in VTE events after the -16.03 percent decrease from 2010 to 2011. In 2012-2013 there was a sharp -33.22 percent decline in VTE events which resulted in the total of a -17.25 percent decline from 2010-2013. The increase from 2011-2012 is not represented in the figure.

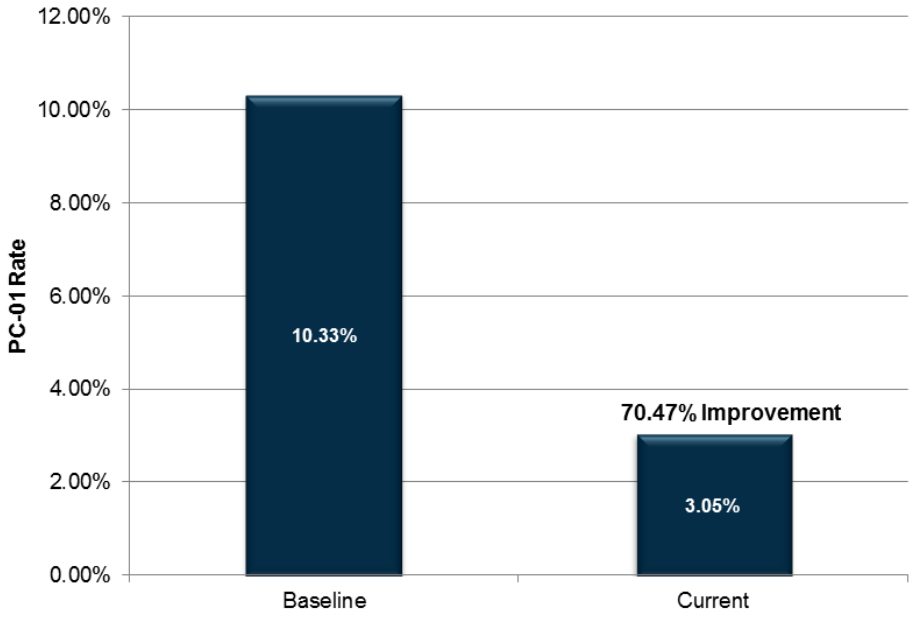
OB-EED

Two measures of EED both show significant improvement. The measure of EED known as PC-01 (The Joint Commission's measure), as self-reported by HENs, shows substantial improvement (70.47 percent, or 7.28 fewer elective deliveries per 100 early deliveries) among participating hospitals. For that measure, the denominator excludes births with 39 weeks or more gestation, so if an OB-EED is prevented, the birth falls out of the denominator. Also, HENs' data contained varying baseline and follow-up periods, so cannot be assessed as a time series, only displayed as a summary bar graph (Figure 2-2). A second measure, which includes all births in the denominator and is calculated from the National Vital Statistics data, shows a 21.00 percent decrease between 2010 and 2013.

Early Elective Delivery (PC-01) (HENs) (Figure 2-2)

- Evidence for improvement.
  - The most recent OB-EED measure is significantly lower than baseline, showing a 70.47 percent improvement (or 7.28 fewer OB-EEDs per 100 early deliveries).

Figure 2-2—Early Elective Delivery (PC-01)

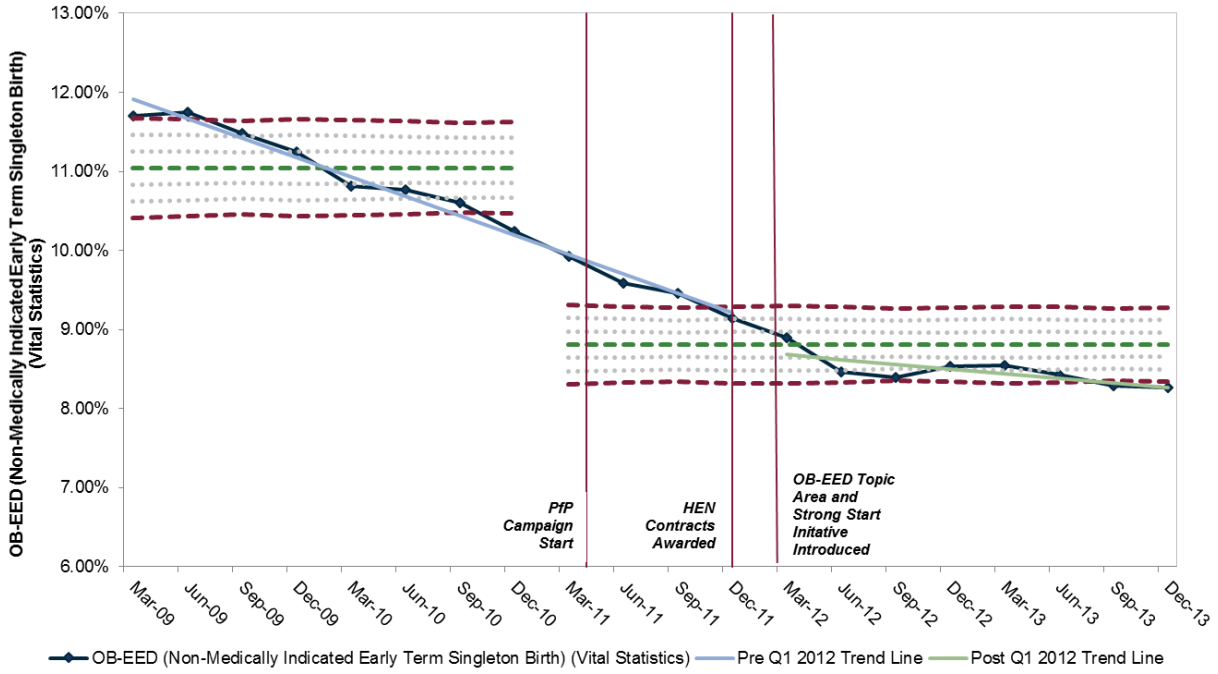


Source: HEN-reported data submitted November 2014.
Note: N = 1,943 hospitals in 25 HENs. Among the 25 HENs reporting on this measure (one of which has four participating cohorts and two of which have two participating cohorts), one baseline period begins in 2010, seven begin in 2011, 10 begin in 2012, and 11 begin in 2013. Most of the baseline periods are quarters, five are annual, one is more than a year, and the remaining use less than a full year but at least 3 months. Twenty-three current periods end in Q1 2014 or later, five end in 2013, and one ends in Q4 2012; one is more than a year and the rest of the current periods are 3 months.

OB-EED (Non-Medically Indicated Early Term Singleton Birth) (Vital Statistics) (Figure 2-3)

- Evidence for improvement.
  - Shift in center line observed in Q1 2011.
- The average quarterly change was a 0.19 percent ($p < 0.001$) lesser decline after Q1 2012 than before.
  - The average rate was shifted 0.47 percent ($p < 0.001$) lower in Q1 2012.
- Rate decreased by 22.07 percent between CY 2010 and Q4 2013.

Figure 2-3—OB-EED (Non-Medically Indicated Early Term Singleton Birth) (Vital Statistics)



Source: National Vital Statistics System (NVSS).
Note: Center line and control limits (U' chart) for first phase were calculated with data between Q1 2009 and Q4 2010. Center line and control limits (U' chart) for second phase were calculated with data between Q1 2011 and Q4 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Readmissions

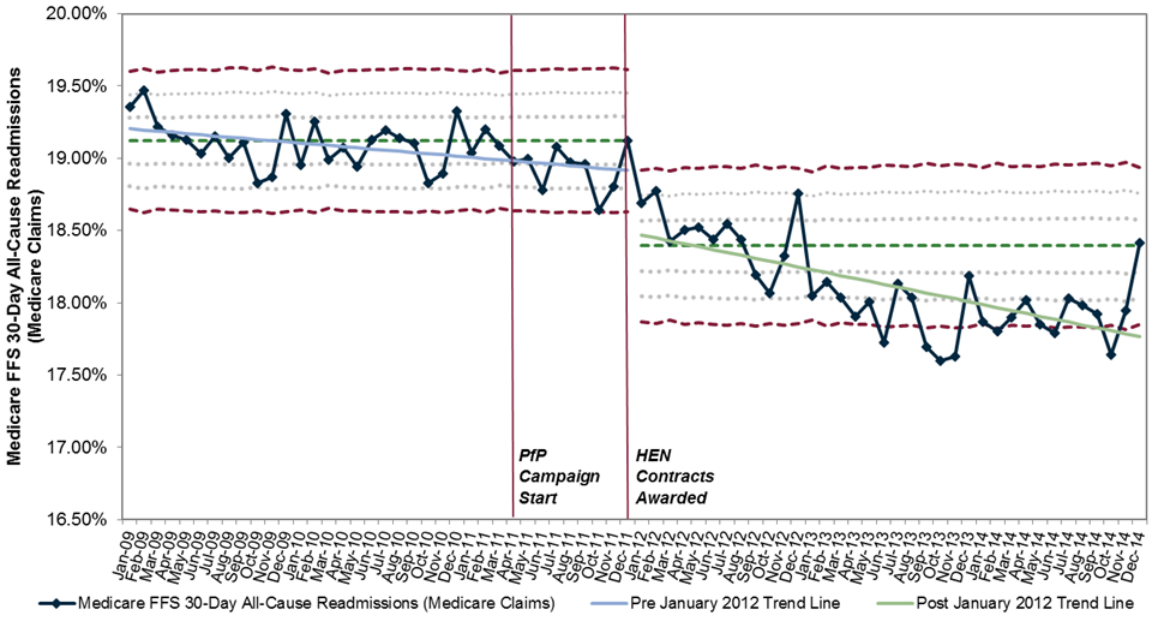
Both the Medicare claims-based measures and the all-payer HEN-reported measures point to improvement in national trends after the start of the PfP campaign. The Medicare claims-based measure, which is complete for short-term acute care hospitals, shows 5.56 percent improvement between CY 2010 and Q4 2014.

Analysis of change in the speed with which the Medicare claims-based rate is changing suggests acceleration of reduction in readmissions in the post-PfP period.

Medicare FFS 30-Day All-Cause Readmissions (Medicare Claims) (Figure 2-4)

- Evidence suggesting improvement in rate.
  - Shift in center line observed in January 2012.
  - Twenty-three data points from January 2013 to November 2014 fall below the second phase center line.
  - Seven data points from June 2013 to October 204 fall below the second phase lower control limit.
- Evidence that the level trended down with a significant break accelerating the rate of decline beginning in January 2012.
  - The average rate is shifted 0.43 percent ($p < 0.05$) lower after January 2012 than before.
  - The monthly rate of decline was 0.01 percent ($p < 0.01$) greater after January 2012 than before.
- Rate decreased 5.56 percent between calendar year 2010 and Q4 2014.

Figure 2-4—FFS 30-Day All-Cause Readmissions (Medicare Claims)

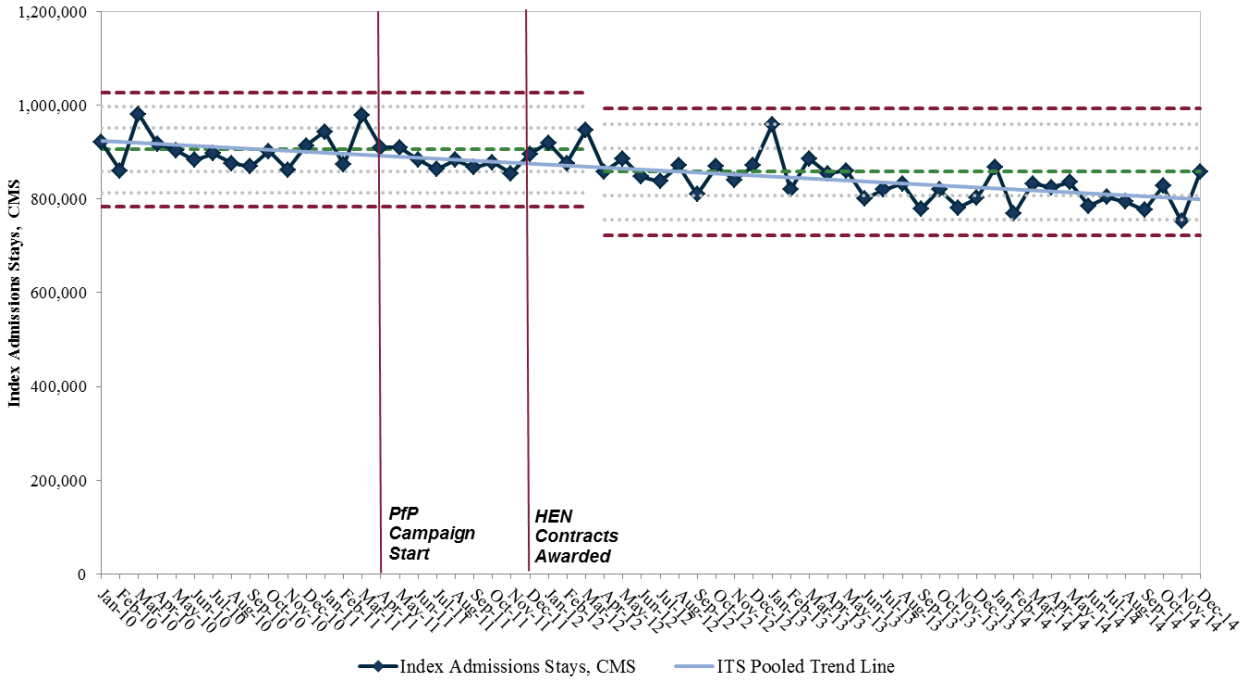


Source: Medicare claims data provided by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS. The Evaluation Contractor processed and ran regression-adjusted analysis to control for changing demographics independently, with similar findings.
Note: Center line and control limits (U' chart) for the first phase were calculated with data between January 2009 and March 2010. Center line and control limits (U' chart) for the second phase were calculated with data between January 2012 and March 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Data include between 981,065 and 754,486 discharges per month.

Index Admissions Stays (Medicare Claims) (Figure 2-5)

- Evidence suggesting improvement in admissions.
  - Shift in center line observed in April 2012.
  - Seventeen data points fall below the second phase center line between June 2013 and December 2014.
- Evidence that admissions trended down without a significant break.
  - The average monthly linear change was -2,133 admissions ($p < 0.05$).
- The number of admissions decreased 11.71 percent between Q1 2010 and Q4 2014.

Figure 2-5—Index Admissions Stays (Medicare Claims)

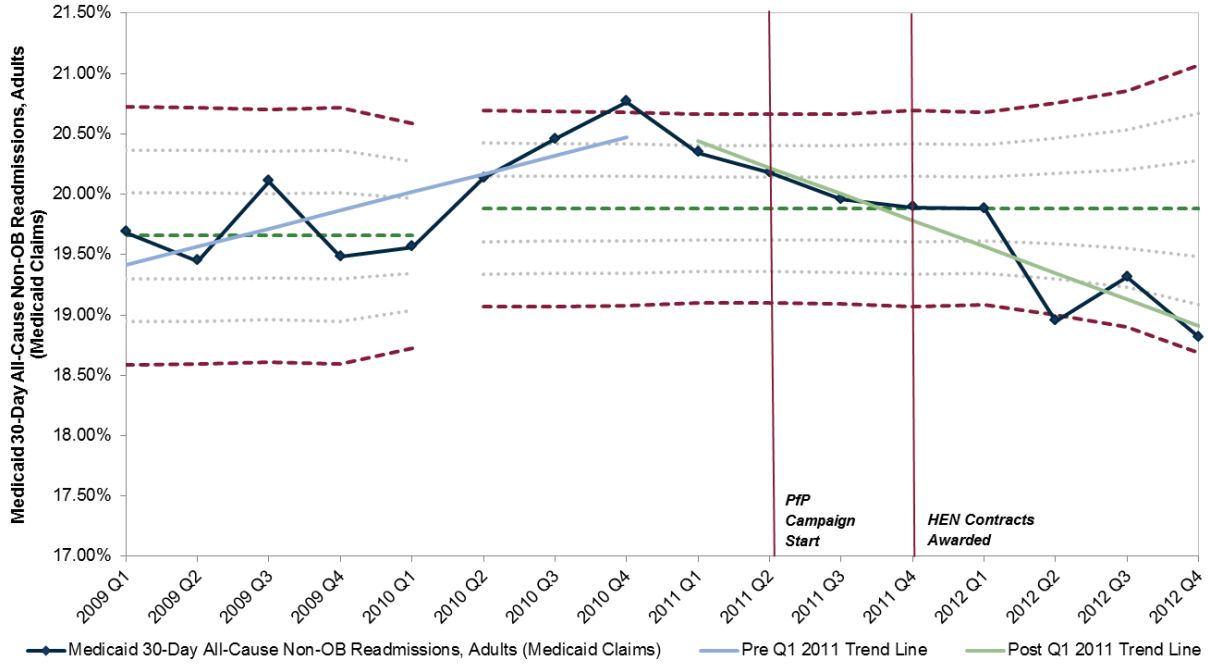


Source: Medicare claims data provided by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS.
Note: Center line and control limits (XmR Chart) for the first phase were calculated with data between January 2010 and March 2011. Center line and control limits (XmR Chart) for the second phase were calculated with data between April 2012 and June 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two sigma limits.

Medicaid 30-Day All-Cause Non-OB Readmissions, Adults (Medicaid Claims) (Figure 2-6)

- Evidence suggesting improvement in rate.
  - Six consecutive data points showing declines from Q1 2011 to Q2 2012.
  - Shift observed in Q2 2010.
- Evidence that the level has improved since Q1 2012.
  - The average quarterly change 0.37 percent ($p < 0.001$) greater decrease after Q1 2011 than before.
- Rate decreased 7.15 percent between CY 2010 and Q4 2012.
- The average quarterly change was a 0.37 percent ($p < .001$) greater decrease after Q1 2011 than before.

Figure 2-6—30-Day All-Cause Non-OB Readmissions, Adults (Medicaid Claims)

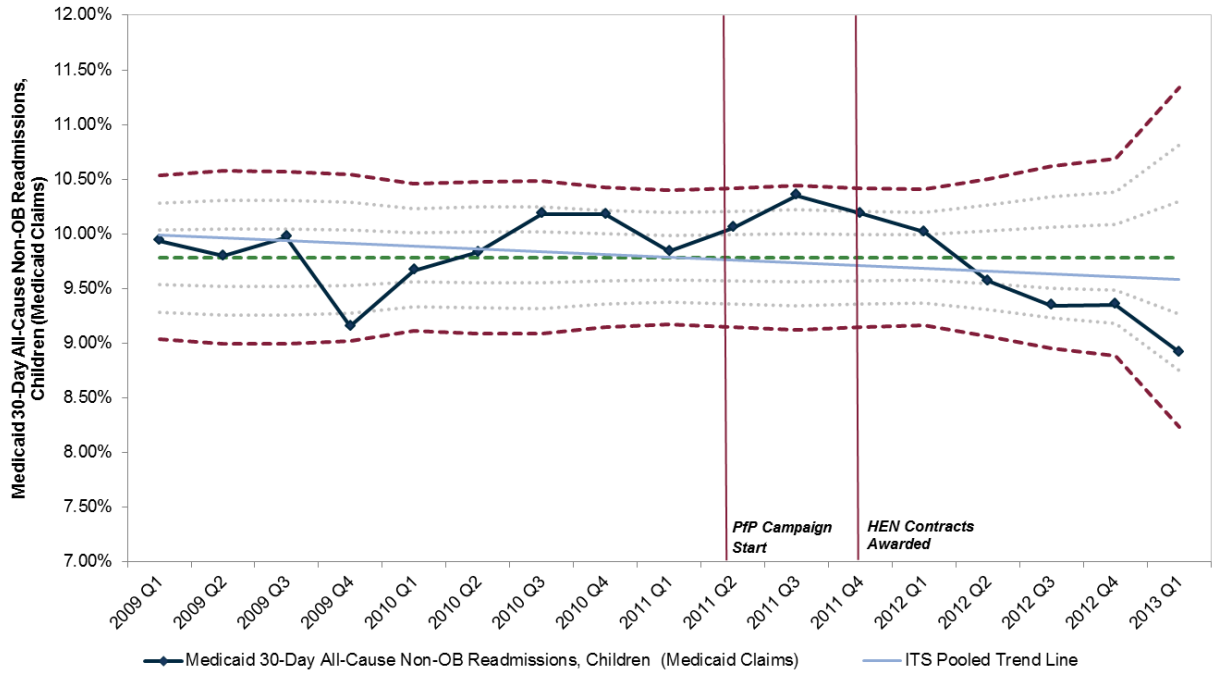


Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q4 2012(Q1 2013 was excluded as an outlier). The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, Medicaid managed care [MMC] and FFS data in 3 states) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 14,920 and 143,489 discharges per quarter. Medicaid data included in these analyses comprise claims for nondual eligible patients age 18 and older.

Medicaid 30-Day All-Cause Non-OB Readmissions, Children (Medicaid Claims) (Figure 2-7)

- No evidence of special cause variation.
- No evidence of a significant break or trend in the rate.
- Rate decreased by 10.50 percent between CY 2010 and Q1 2013.

Figure 2-7—30-Day All-Cause Non-OB Readmissions, Children (Medicaid Claims)



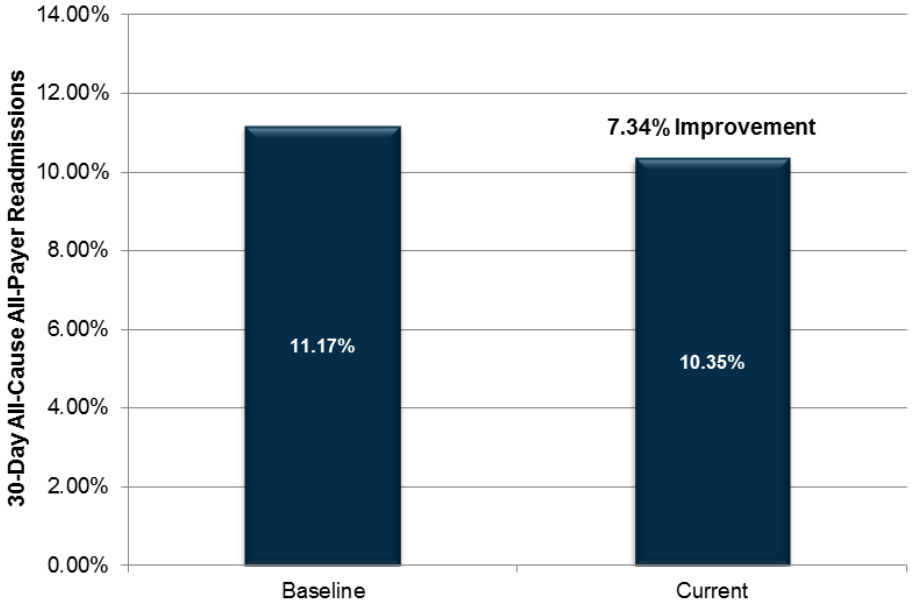
Source: Medicaid claims data for 17 states.

Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q1 2013. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states), assuming all diagnosis codes were not POA (Medicaid claims do not include POA data) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 5,698 and 36,508 discharges per quarter. Medicaid data included in these analyses comprise claims for nondual eligible patients ages 0 to 17.

30-Day All-Cause All-Payer Readmissions (HENs) (Figure 2-8)

- Evidence for improvement from the comparison of rates (*z*-test).
  - A statistically significant 7.34 percent improvement (0.82 percentage points) in the current period compared to the baseline period.

Figure 2-8—30-Day All-Cause All-Payer Readmissions (HENs)



Source: HEN-reported data submitted November 2014. HENs' measure definitions may vary.
Note: N = 2,517 hospitals in 22 HENs. Baseline and current period rates are rounded for presentation, while the percentage improvement is calculated using unrounded data. HEN baselines and current periods vary. Among 22 HENs reporting on this measure (one of which has three participating cohorts and three of which have two participating cohorts), 15 baseline periods began in 2010, five began in 2011, and four began in 2012. Twelve baseline periods are annual, 10 are at least 3 months but less than a year, and two are more than a year. Eighteen current periods end in Q1 2014 or later, five end in 2013, and one ends in 2012; one current period is 6 months and the rest are 3 months.

All-Payer, All-Cause, All Diagnosis 30-Day Readmissions per 100 Index Admissions (AHRQ) (Table 2-2)

- Evidence for improvement, with a small decrease (2.6 percent) in rates from 2010 to 2013.

Table 2-2—All-Payer, All-Cause, All-Diagnosis 30-Day Readmissions per 100 Index Admissions (AHRQ)

2010 Rate (per 100 index admissions)	2011 Rate (per 100 index admissions)	2012 Rate (per 100 index admissions)	2013 Rate (per 100 index admissions)
14.40	14.42	14.25	14.03

Source: Readmissions Data Summary for PfP Use, Data from Joanna Jiang of AHRQ CDOM, updated 21 August 2015.
Note: The estimates incorporate HCUP data analyzed by AHRQ, and Medicare readmission rates provided by CMS. The 2013 HCUP data used for the table (received on August 13) is not public yet and the public final version may diverge slightly from this version.

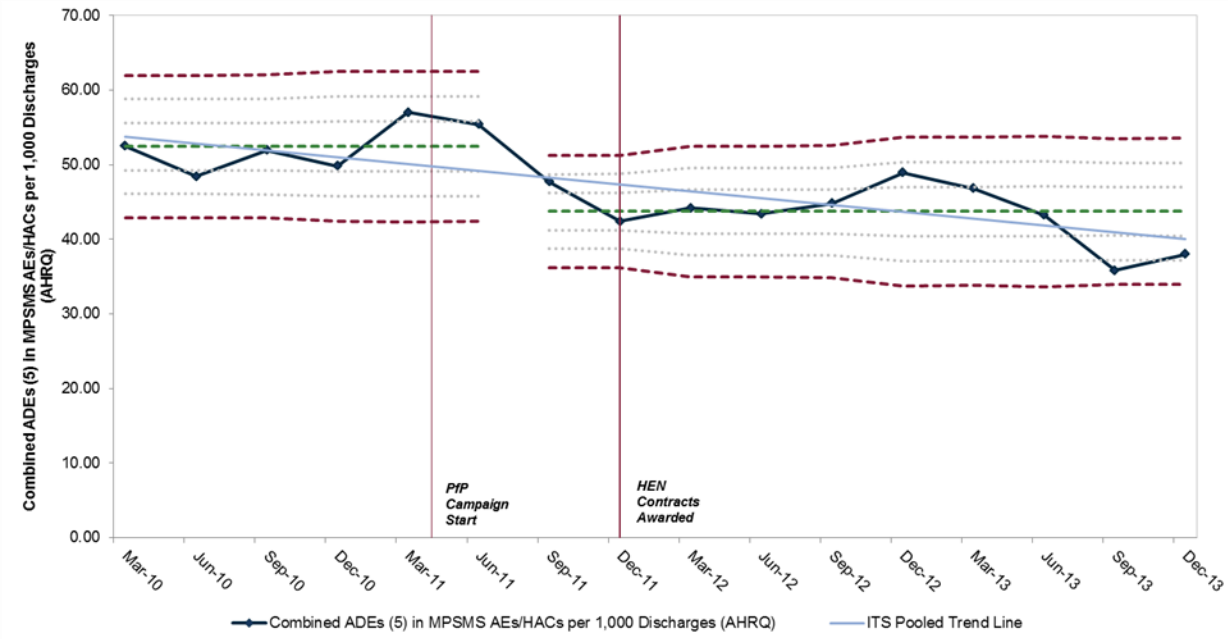
ADE

The only available measure for ADEs, from the AHRQ National Scorecard shows substantial improvement for both its quarterly and annual data (25.00 percent and 18.58 percent) in national trends. It should be noted that about 99 percent of AHRQ's ADE data are on only two major sources of ADEs, namely, (1) ADEs associated with the use of insulin and other hypoglycemic agents and (2), ADEs associated with anticoagulants.[2-4]

Combined ADEs (5) in MPSMS AEs/HACs per 1,000 Discharges (AHRQ National Scorecard, Quarterly) (Figure 2-9)[2-5]

- Evidence suggesting improvement in rate.
  - Shift in center line observed in Q3 2011.
- Evidence that the rate trended downward without a significant break.
  - The average quarterly linear changes was -0.92 per 1,000 discharges ($p < 0.01$).
- Rate decreased by 25.00 percent between CY 2010 and Q4 2013.

Figure 2-9—Combined ADEs (5) in MPSMS AEs/HACs per 1,000 Discharges (AHRQ, Quarterly)



Source: AHRQ National Scorecard (quarterly data).
Note: Center line and control limits (U' chart) for first phase were calculated with data between Q1 2010 and Q2 2011. Center line and control limits (U chart) for second phase were calculated with data between Q3 2011 and Q4 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

2-4 See http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html.
2-5 The Evaluation Contractor requested AHRQ provide quarterly data for ADEs to produce an SPC chart. The AHRQ National Scorecard produces the only measure for ADEs.

ADEs per 1,000 Discharges (AHRQ National Scorecard, Annual) (Table 2-3)

- Evidence for improvement, with the rate for 2013 being 18.58 percent lower than the rate in 2010.

Table 2-3—ADEs per 1,000 Discharges (AHRQ, Annual)

2010 HAC Rates (per 1,000 discharges)	2011 HAC Rates (per 1,000 discharges)	2012 HAC Rates (per 1,000 discharges)	2013 HAC Rates (per 1,000 discharges)
49.50	48.67	41.91	40.30

Source: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html and Noel Eldridge, AHRQ Center for Quality Improvement and Safety, provided on August 28, 2015, for the final 2013 data.

VTE

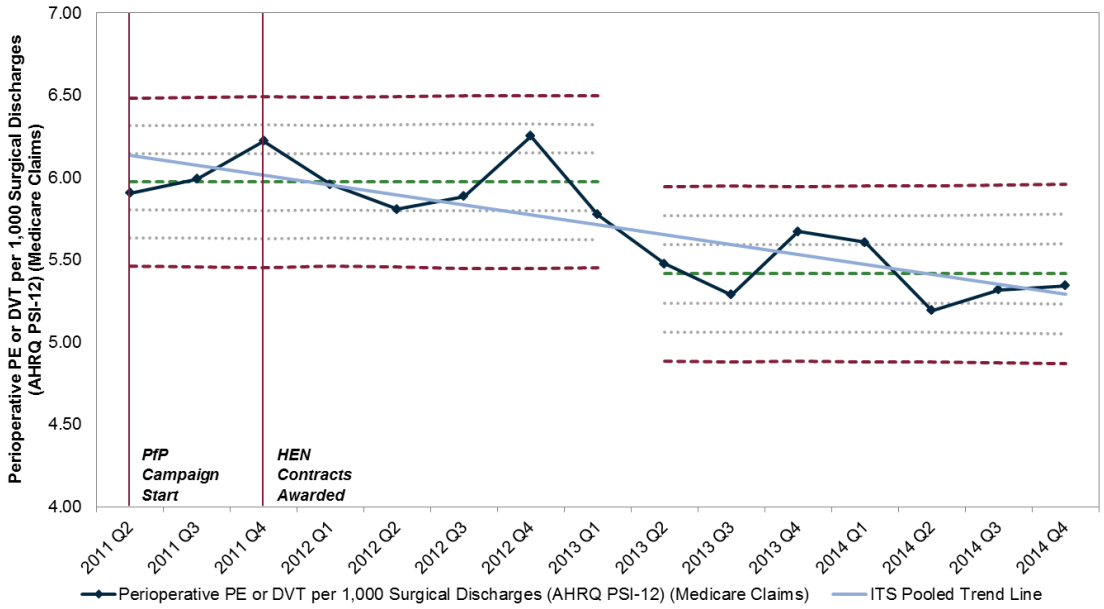
The analysis of trends in post-operative VTE depended on the source of data and the test used. All-payer data from two sources show a decrease: AHRQ National Scorecard data (16.47 percent decrease), and HEN-submitted all-payer data (10.71 percent decrease). Medicare data showed a decrease that began in Q4 2012 and was sustained through Q4 2014. The Medicaid data did not show a significant decrease through Q4 2012.

Analyses of change in the speed with which the Medicare and Medicaid claims-based rates are changing suggests no change in the post-PfP period.

Perioperative PE or DVT per 1,000 Surgical Discharges (AHRQ PSI-12) (Medicare Claims) (Figure 2-10)

- Evidence suggesting improvement in rate.
  - Shift in center line observed in Q2 2013.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.06 per 1,000 surgical discharges ($p < 0.05$).
- Rate decreased by 9.55 percent between Q2 2011 and Q4 2014.

Figure 2-10—Perioperative PE or DVT per 1,000 Surgical Discharges (Medicare Claims)

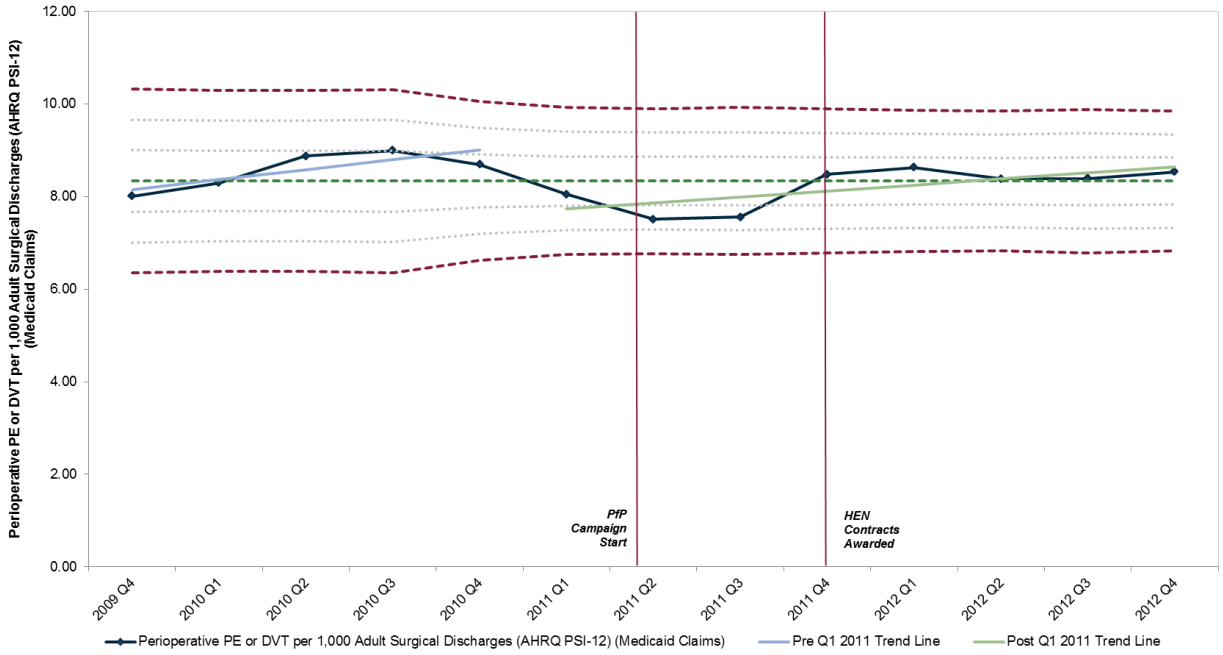


Source: Rates calculated by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS based on Medicare FFS claims data. The Evaluation Contractor also conducted a similar analysis, adding a longer time series (from 2009) and using regression analysis to control for changing demographics over time. The data showed a similar pattern.
Note: Control limits and center line (U′ chart) for the first phase were calculated with data between Q2 2011 to Q1 2013.Control limits and center line (U' chart) for the second phase were calculated with data between Q2 2013 to Q4 2014.The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma lines; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicare FFS claims for all hospitals that reliably report POA status (≥95 percent of the hospital's diagnoses for a given quarter are accompanied by a valid code for POA) and that have the following characteristics: all hospitals paid under Medicare's inpatient prospective payment system (IPPS), critical access hospital (CAH), cancer hospitals, and Maryland hospitals. Data include between 619,182 and 690,101 discharges per quarter.

Perioperative PE or DVT per 1,000 Adult Surgical Discharges (AHRQ PSI-12) (Medicaid Claims) (Figure 2-11)

- No evidence of special cause variation.
- Evidence of a significant break or trend in the rate.
  - The average rate was shifted 1.14 ($p < 0.01$) lower in Q1 2011.
- Rate increased by 2.09 percent between CY 2010 and Q4 2012.
- This measure may contain community-acquired as well as hospital-acquired PE or DVT, since Medicaid claims do not contain information about whether the condition was present on admission.

Figure 2-11—Perioperative PE or DVT per 1,000 Adult Surgical Discharges (Medicaid Claims)



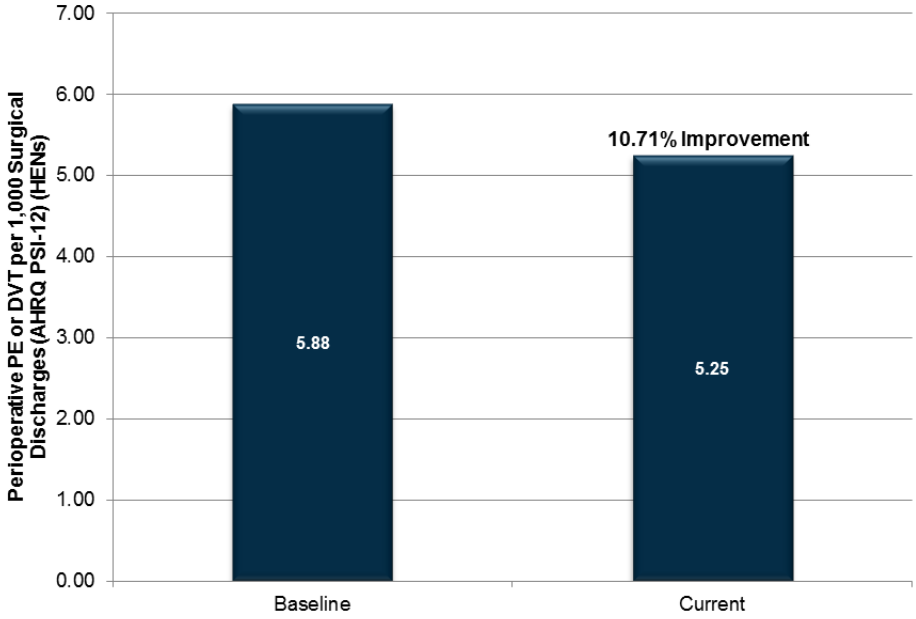
Source: Medicaid claims data for 17 states.

Note: Control limits and center line (U chart) constructed using data from Q4 2009 to Q4 2012. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states), assuming all diagnosis codes were not POA (Medicaid claims do not include POA data) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 18,923 and 33,159 discharges per quarter. Medicaid data included in these analyses comprise claims for nondual eligible patients age 18 and older.

Perioperative PE or DVT per 1,000 Surgical Discharges (AHRQ PSI-12) (HENs) (Figure 2-12)

- Evidence for improvement.
  - The most recent PSI-12 measure is significantly lower than baseline, showing a 10.71 percent improvement.

Figure 2-12—Perioperative PE or DVT per 1,000 Surgical Discharges (HENs)



Source: HEN-reported data submitted November 2014.
Note: N = 2,282 hospitals in 22 HENs. Among the 22 HENs reporting on this measure (one of which has two participating cohorts), 13 baseline periods begin in 2010; seven begin in 2011, and three begin in 2012. Ten of the baseline periods are annual, eight are 3 months, two are 6 months, and three are more than a year. Nineteen current periods end in Q1 2014 or later and four end in 2013; most current periods are 3 months, one is 6 months, one is a year, and one is more than a year.

Post-Operative VTE per 1,000 Discharges (AHRQ National Scorecard) (Table 2-4)

- Evidence for improvement.
  - The rate for 2013 is 16.47 percent lower than the rate in 2010. It should be noted, however, that this rate varies highly from year to year, as it is based on a small number of harms.

Table 2-4—Post-Operative VTE per 1,000 Discharges* (AHRQ)

2010 HAC Rates (per 1,000 discharges)	2011 HAC Rates (per 1,000 discharges)	2012 HAC Rates (per 1,000 discharges)	2013 HAC Rates (per 1,000 discharges)
0.85	0.72	0.99	0.71

Source: Noel Eldridge, AHRQ Center for Quality Improvement and Safety, provided on August 28, 2015 (for the final 2013 data) and http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html.

*Although national data on number of surgeries were not available to the Evaluation Contractor at the time of this report, if the total number of surgeries trended the same way as the number of surgeries covering 17 common procedures (which is available), then if a surgical discharges number were used for the denominator rather than a total discharges number, the decrease could be closer to 18 percent.

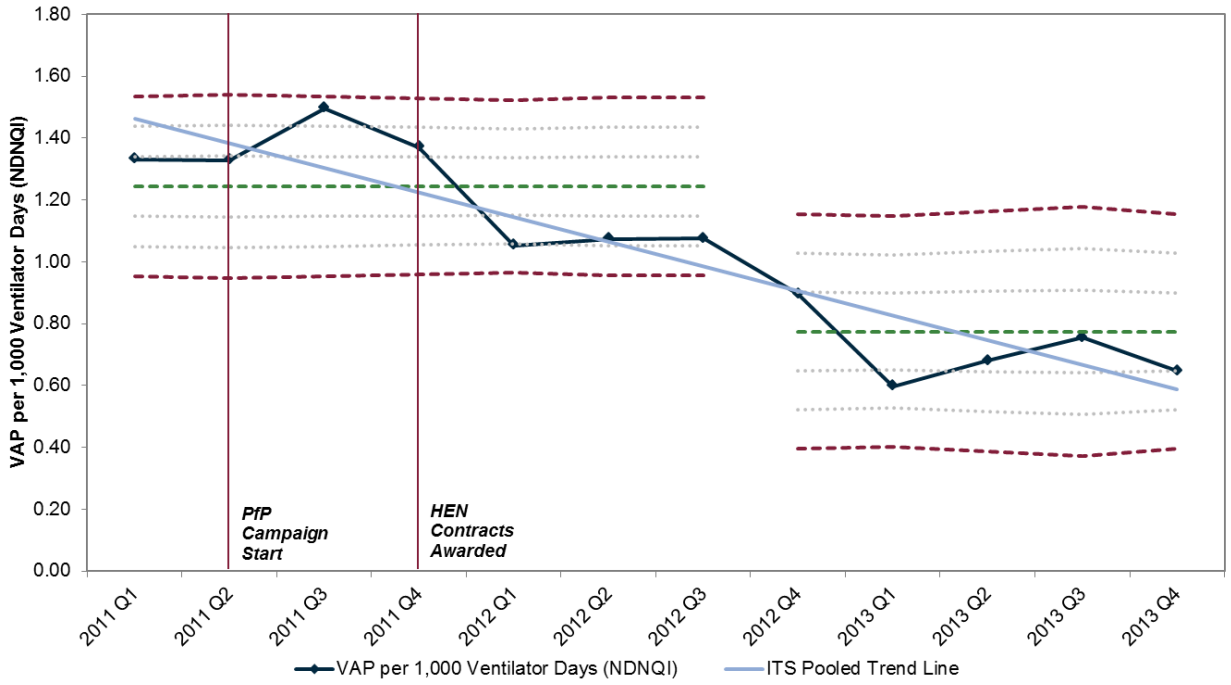
VAP/VAE

The NDNQI measure of VAP per 1,000 ventilator days, including the comparison of rates in this measure, and the AHRQ National Scorecard measure all indicate improvement in national trends: a 3.45 percent reduction between 2010 and 2013 in the AHRQ National Scorecard, which are the most complete data source (Table 2-5, below).

VAP per 1,000 Ventilator Days (NDNQI) (Figure 2-13)[2-6]

- Evidence suggesting improvement in rate.
  - Shift in center line observed in Q4 2012.
  - Control limits are temporary pending additional data.[2-7]
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.08 per 1,000 ventilator days ($p < 0.05$).
- Rate decreased 51.42 percent between Q1 2011 and Q4 2013.

Figure 2-13—VAP per 1,000 Ventilator Days (NDNQI)



Source: NDNQI. Data are between 484 and 544 hospitals per quarter.
Note: Control limits and center line (U' chart) for first phase constructed using data from Q1 2011 to Q3 2012. Control limits and center line (U' chart) for second phase constructed using data from Q4 2012 to Q4 2013. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. VAP data after Q4 2013 are not comparable due to a change in definition and a large reduction in the number of reporting hospitals.

2-6 The VAP measure definition used has been changed by the CDC due to concerns about variable interpretation; however, no other sufficiently broad data are available for analysis.

2-7 For SPC charts with fewer than 16 observations, the control limits are unstable, and thus final control limits cannot be calculated without additional data.

VAP per 1,000 Discharges (AHRQ National Scorecard) (Table 2-5)

- Evidence for improvement.
  - The rates from the National Scorecard show a decline from 2010 to 2012. The 2012 rate is 12.07 percent lower than the 2010 rate (0.14 fewer VAPs per 1,000 discharges). However, the 2013 rate increased to 1.12 per 1,000 discharges, resulting in a net reduction from 2010 of 3.45 percent.

Table 2-5—VAP per 1,000 Discharges (AHRQ)

2010 HAC Rates (per 1,000 discharges)	2011 HAC Rates (per 1,000 discharges)	2012 HAC Rates (per 1,000 discharges)	2013 HAC Rates (per 1,000 discharges)
1.16	1.07	1.02	1.12

Source: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html and Noel Eldridge, AHRQ Center for Quality Improvement and Safety, provided on August 28, 2015, for the final 2013 data.

CAUTI

Whether there is an improvement in national trends for CAUTI results depends upon the data source and the rate comparison methodology used. Among all-payer data, the AHRQ National Scorecard data, which are most complete, show the 2013 rate 27.52 percent lower than 2010. However, the CAUTI SIR for ICUs from the NHSN data appears to increase between Q1 2012 and Q1 2014. In contrast, the CAUTI SIR and device utilization rate for non-ICUs in the NHSN data exhibit significant downward trends, indicating that both the use of urinary catheters and infection rates associated with urinary catheter use were declining outside the ICU. Thus, the reduction in CAUTI rates in the NHSN data suggest a complex relationship wherein rates may be reduced among non-ICU patients in part through a reduction in the reliance on urinary catheters.

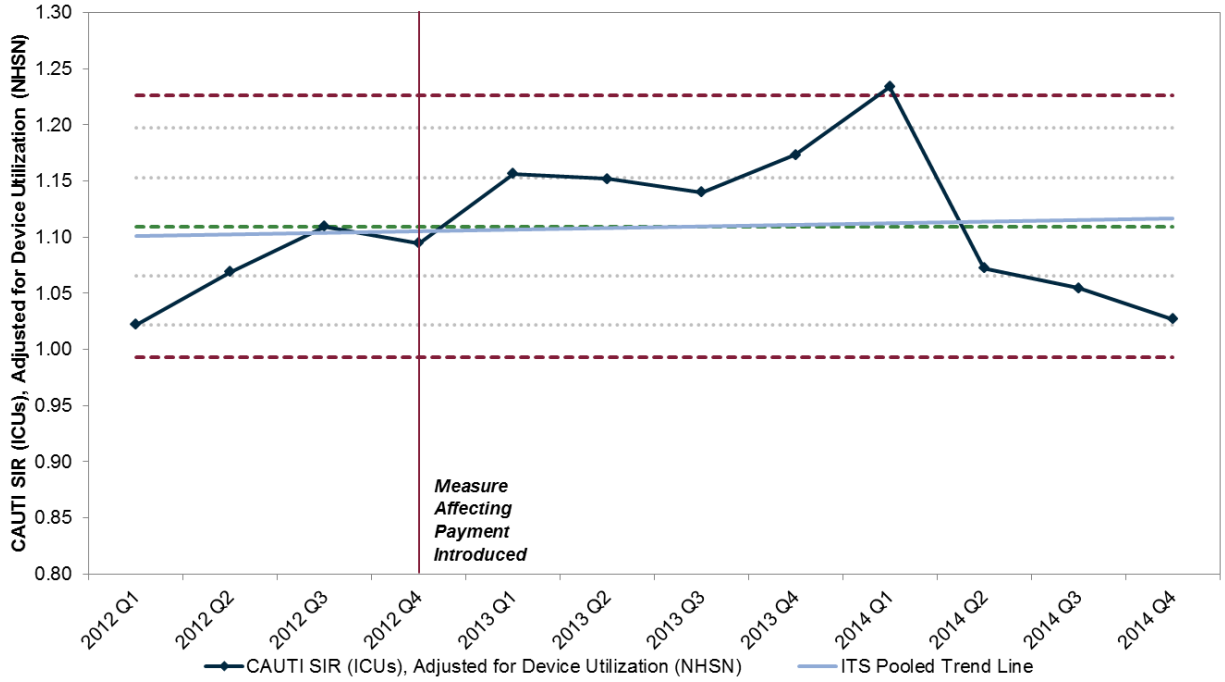
A measure with limited reporting, CAUTI per 1,000 catheter days (NDNQI), shows substantial improvement between baseline (Q1 2011) and the post-HEN startup period (Q1 2012); however, the rate has been steady in this source in recent quarters, and the most recent quarter is 16.53 percent below baseline. Analyses of Medicaid claims show improvement on most comparisons made. The exception is that Medicaid analyses of HAUTI rates among adult populations exhibit different results depending on the test; the control chart shows a signal of no change but the pre/post comparison testing shows significant change in trend from baseline, and a 2.79 percent worsening of the rate. Note that trends in HAUTI are included in this section as well as those infections that are coded in administrative data as CAUTI; this is due to the knowledge that CAUTI is under-coded in administrative data; however, the HAUTI rate will include infections that are both catheter-related and non-catheter-related. Therefore, the HAUTI rate will be higher than the rate that would be observed for CAUTI alone.

Analysis of change in the speed with which the CAUTI rate is changing in the post-PfP period compared to the baseline period suggests worsening in the all-payer NDNQI measure and one Medicaid claims-based trend, and no change in the NHSN trends.

CAUTI SIR (ICUs), Adjusted for Device Utilization (NHSN) (Figure 2-14)

- Evidence of special cause variation.
  - Q1 2014 falls above the upper control limit.
  - Center line and control limits are temporary pending additional data.[2-8]
- No evidence of significant break or trend in the rate.
- Rate increased by 0.47 percent between Q1 2012 and Q4 2014.

Figure 2-14—CAUTI SIRs (ICUs), Adjusted for Device Utilization (NHSN)



Source: NHSN. Data are between 3,072 and 3,150 hospitals per quarter. Data are adjusted for change in utilization, which also changes risk of harm. To adjust, the national unadjusted CAUTI SIR (ICUs) data points are multiplied by the ratio of the national device utilization rate for that period to the national device utilization rate in the first quarter.

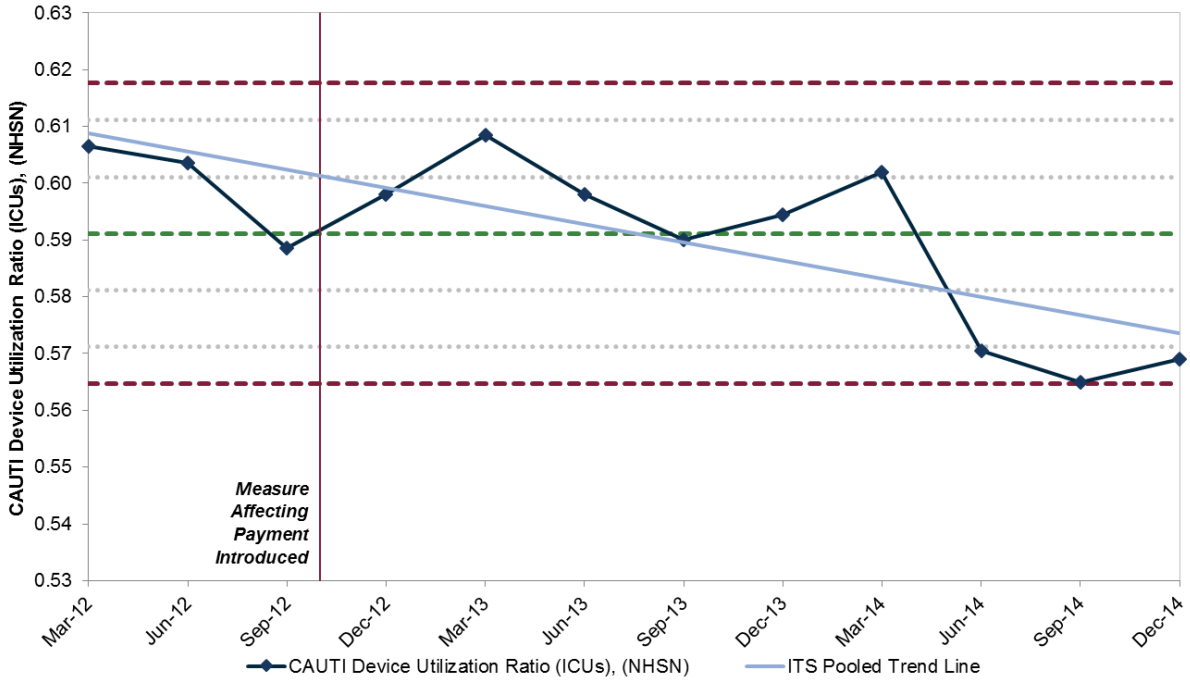
Note: Control limits and center line (X chart) are constructed using data from Q1 2012 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. The data from NHSN were not provided with clustered standard errors that would account for the non-independence of discharges within the same facilities. Without the clustered standard errors, the SPC charts and ITS analysis performed are less conservative than would be the case if the correlation of outcomes within facilities were accounted for in the analysis.

[2-8] For SPC charts with fewer than 16 observations, the control limits are unstable, and thus final control limits cannot be calculated without additional data.

CAUTI Device Utilization Ratio (ICUs), (NHSN) (Figure 2-15)

- Evidence of special cause variation.
  - Three data points from Q2 2014 to Q4 2014 fall between the -2σ limit and the lower control limit.
- The average quarterly linear change was 0.32 percent ($p < 0.05$).
- The device utilization ratio decreased 6.18 percent between Q1 2011 and Q4 2014.

Figure 2-15—CAUTI Device Utilization Ratio (ICUs), (NHSN)

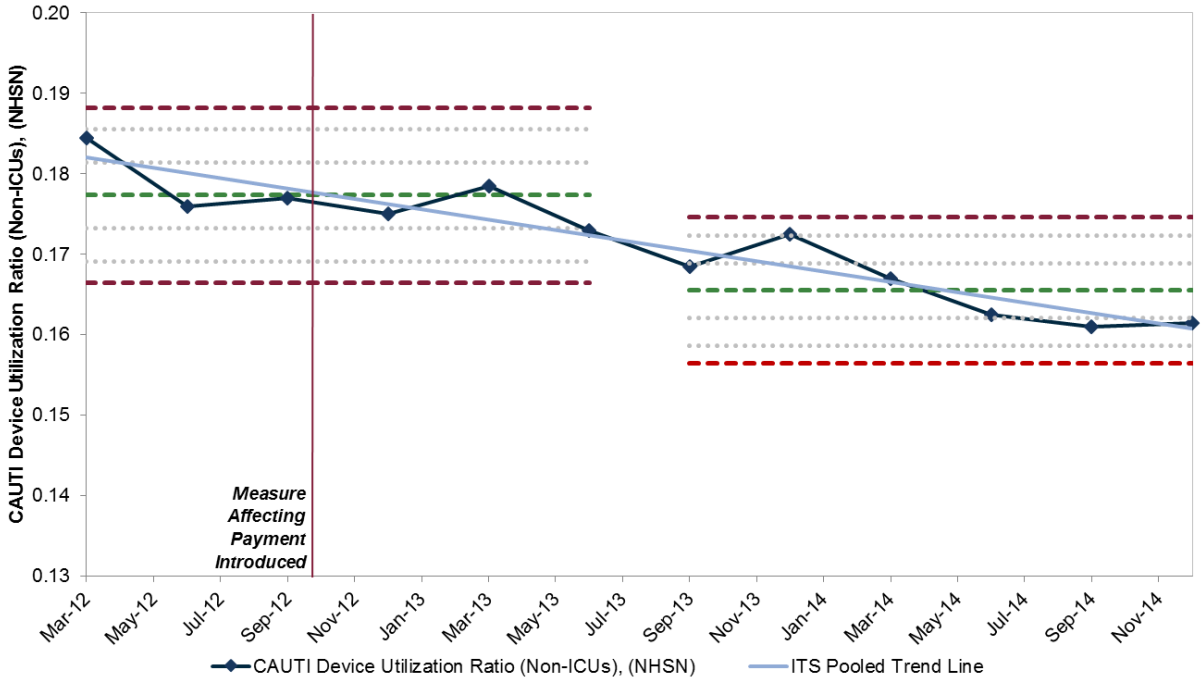


Source: NHSN. Data are between 3,072 and 3,150 hospitals per quarter.
Note: Center line and control limits (XmR chart) were calculated with data between Q1 2012 and Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two sigma limits.

CAUTI Device Utilization Ratio (Non-ICUs), (NHSN) (Figure 2-16)

- Evidence suggesting improvement in rate.
  - Shift in center line observed in Q3 2013.
- The average quarterly linear change was 0.19 percent ($p < 0.001$).
- The device utilization ratio decreased 12.47 percent between Q1 2011 and Q4 2014.

Figure 2-16—CAUTI Device Utilization Ratio (Non-ICUs), (NHSN)

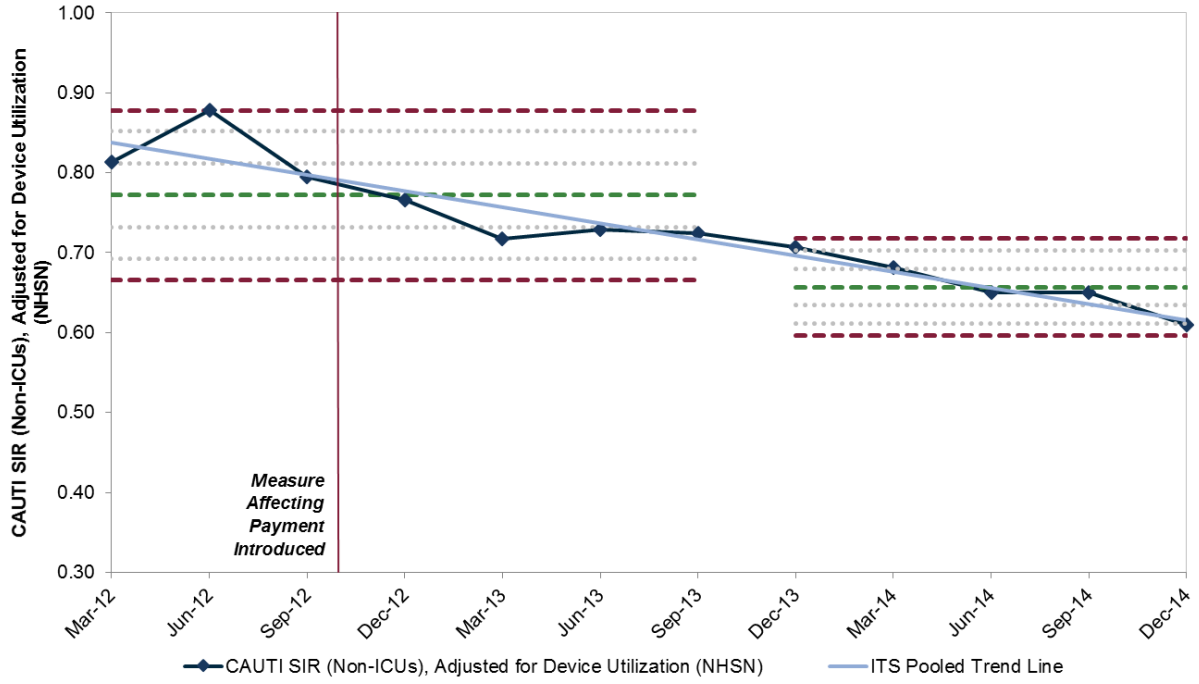


Source: NHSN. Data are between 1,197 and 1,813 hospitals per quarter.
Note: Center line and control limits (XmR Chart) for first phase were calculated with data between Q1 2012 and Q2 2013. Center line and control limits (XmR Chart) for first phase were calculated with data between Q3 2013 and Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two sigma limits.

CAUTI SIR (Non-ICUs), Adjusted for Device Utilization (NHSN) (Figure 2-17)

- Evidence suggesting improvement in rate.
  - Shift in center line observed in Q4 2013.
  - Six consecutive decreases in rate from Q2 2013 to Q4 2014.
- The average quarterly linear change was 2.03 percent ($p < 0.001$).
- The SIR decreased 24.97 percent between Q1 2011 and Q4 2014.

Figure 2-17—CAUTI SIR (Non-ICUs), Adjusted for Device Utilization (NHSN)

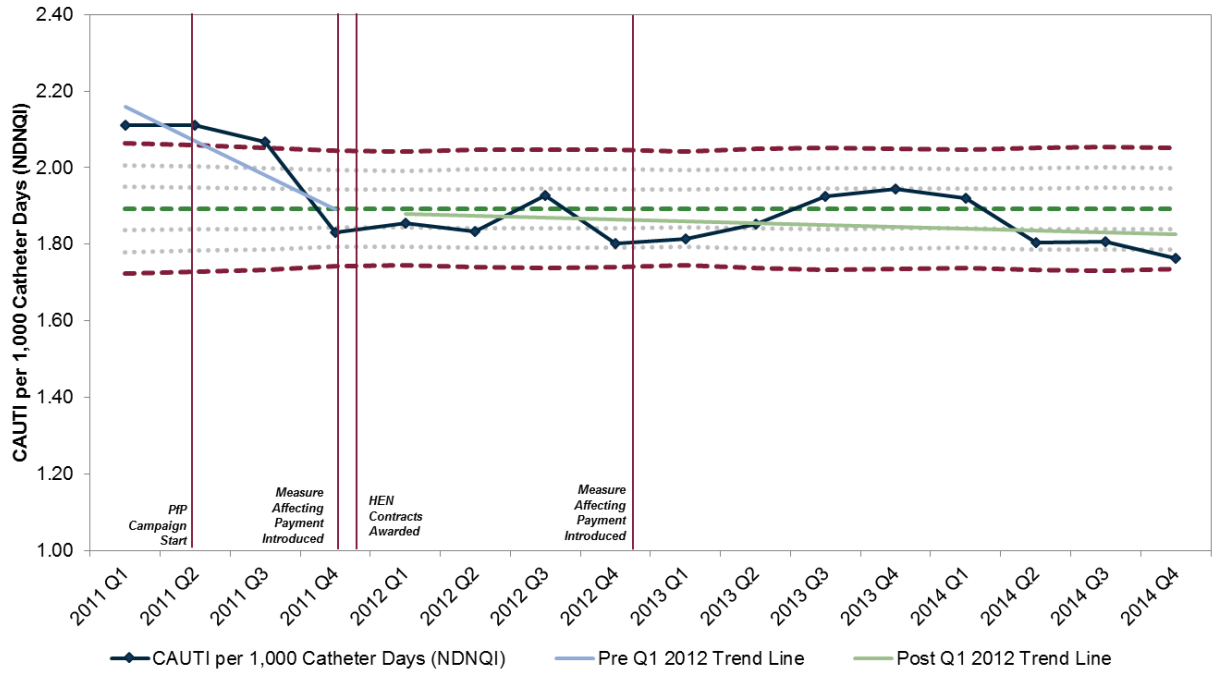


Source: NHSN. Data are between 1,197 and 1,813 hospitals per quarter. Data are adjusted for change in utilization, which also changes risk of harm. To adjust, the national unadjusted CAUTI SIR (Non-ICUs) data points are multiplied by the ratio of the national device utilization.
Note: Center line and control limits (XmR Chart) for first phase were calculated with data between Q1 2012 and Q3 2013. Center line and control limits (XmR Chart) for first phase were calculated with data between Q4 2013 and December 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two sigma limits.

CAUTI per 1,000 Catheter Days, All Tracked Units (NDNQI) (Figure 2-18)

- No recent evidence to suggest special cause variation.
  - First three data points are above the upper control limit.
- Evidence suggesting that the trend has worsened since Q1 2012.
  - The average quarterly change was a 0.08 per 1,000 catheter days ($p < 0.05$) lesser decline after Q1 2012 than before.
- Rate decreased by 16.53 percent between Q1 2011 and Q4 2014.
- Inpatient Prospective Payment System (IPPS) payment incentive related to CAUTIs became effective October 2011; CMS stopped paying for the HAC for CAUTI October 2012.

Figure 2-18—CAUTI per 1,000 Catheter Days (NDNQI)

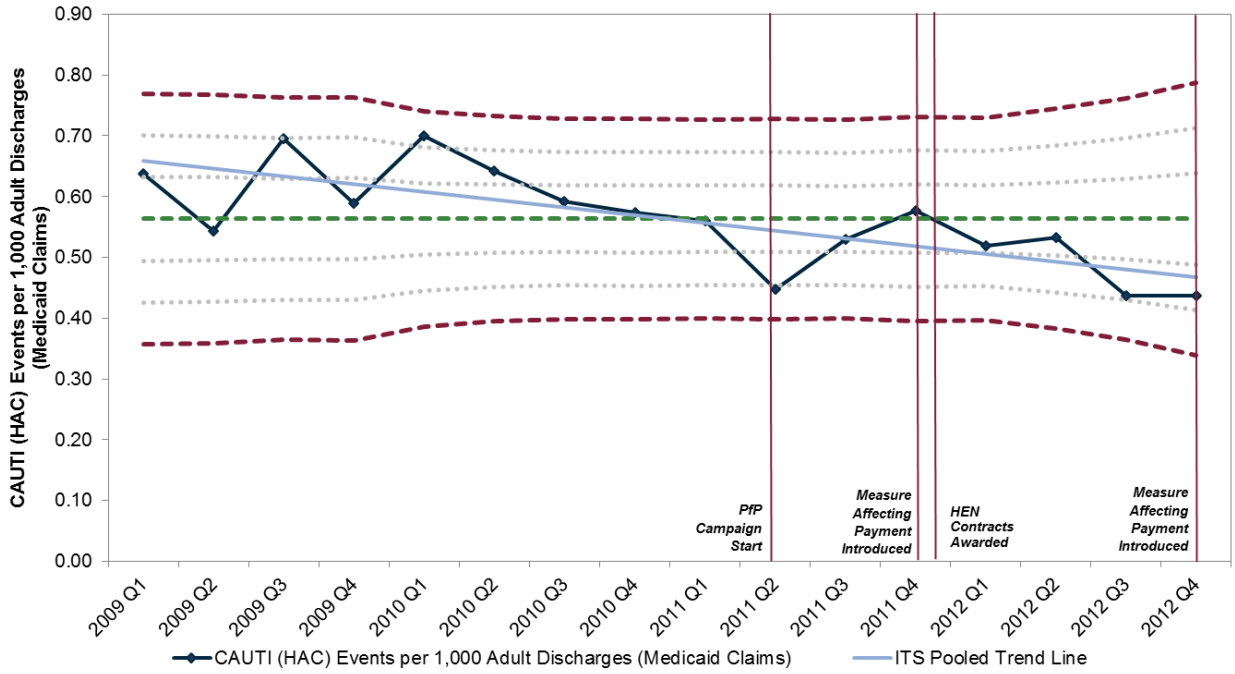


Source: NDNQI. Data are between 554 and 680 hospitals per quarter.
Note: Control limits and center line (U′ chart) were constructed using data from Q1 2011 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

CAUTI (HAC) Events per 1,000 Adult Discharges (Medicaid Claims) (Figure 2-19)

- No evidence to suggest special cause variation.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.013 per 1,000 adult discharges ($p < 0.001$).
- The rate decreased by 30.19 percent between CY 2010 and Q4 2012.
- IPPS payment incentive related to CAUTIs became effective October 2011; CMS HAC for CAUTI became effective October 2012.
- This measure includes community-acquired as well as hospital-acquired CAUTI, since Medicaid claims do not contain data for whether the condition was present on admission.

Figure 2-19—CAUTI (HAC) Events per 1,000 Adult Discharges (Medicaid Claims)



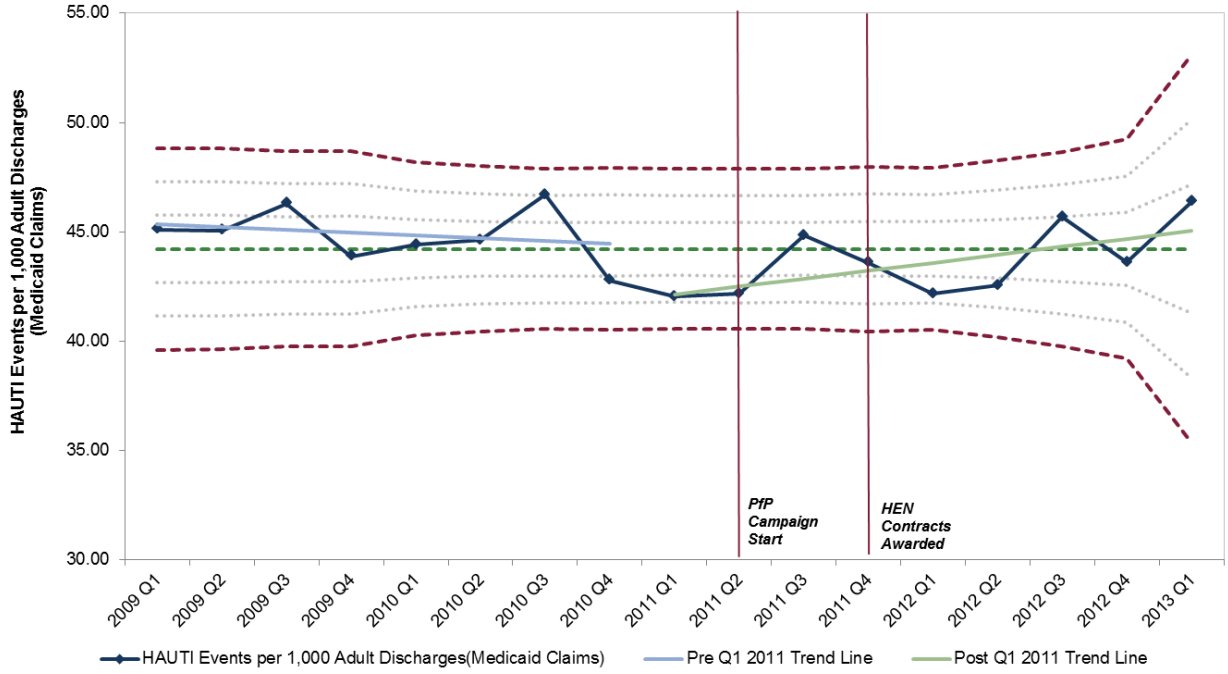
Source: Medicaid claims data for 17 states.

Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q4 2012. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states; MMC and FFS data in 3 states), assuming all diagnosis codes were not POA (Medicaid claims do not include POA data) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 129,768 and 246,083 discharges per quarter. Medicaid data included in these analyses comprise claims for nondual eligible patients age 18 and older.

HAUTI Events per 1,000 Adult Discharges (Medicaid Claims) (Figure 2-20)[2-9]

- No evidence to suggest special cause variation.
- Evidence that the measure has worsened since Q1 2011.
  - The average quarterly change was a 0.49 ($p < 0.05$) per 1,000 greater increase after Q1 2011 than before.
  - The average rate was shifted 2.71 ($p < 0.05$) lower in Q1 2011.
- Rate increased 3.89 percent between CY 2010 and Q1 2013.
- This measure includes community-acquired as well as hospital-acquired HAUTI, since Medicaid claims do not contain data for whether conditions were present on admission.

Figure 2-20—HAUTI Events per 1,000 Adult Discharges (Medicaid Claims)



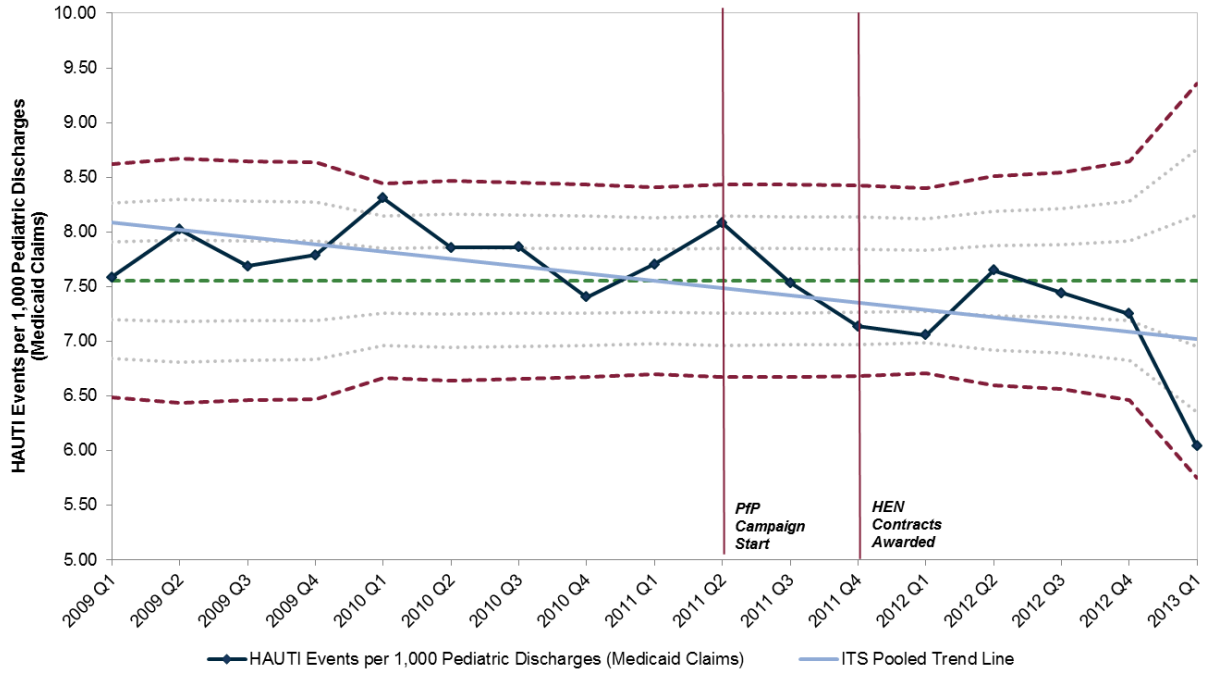
Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q1 2013. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states), assuming all diagnosis codes were not POA (Medicaid claims do not include POA data) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 41,630 and 246,083 discharges per quarter. Medicaid data included in these analyses comprise claims for nondual eligible patients age 18 and older.

2-9 Similar to Medicaid data, Medicare data also lacks the POA indicators. As the HAUTI measure numerator includes a much more general code for UTI, 599.0 (urinary tract infection, site not specified) than CAUTI, the lack of POA indicators likely also has a much larger effect on HAUTI rates versus CAUTI rates.

HAUTI Events per 1,000 Pediatric Discharges (Medicaid Claims) (Figure 2-21)

- No evidence to suggest special cause variation.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.067 per 1,000 pediatric discharges ($p < 0.01$).
- Rate decreased 23.13 percent between CY 2010 and Q1 2013.
- This measure includes community-acquired as well as hospital-acquired HAUTI, since Medicaid claims do not contain data for whether the condition was present on admission.

Figure 2-21—HAUTI Events per 1,000 Pediatric Discharges (Medicaid Claims)



Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q1 2013. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states), assuming all diagnosis codes were not POA (Medicaid claims do not include POA data) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 27,527 and 124,421 discharges per quarter. Medicaid data included in these analyses comprise claims for nondual eligible patients ages 0 to 17.

CAUTI per 1,000 Discharges (AHRQ National Scorecard) (Table 2-6)

- Evidence for improvement from raw rates in this measure.
  - Steady improvement in rates (falling rates of harms) in 2010, 2011, 2012, and 2013, with the 2013 rate 27.52 percent lower than the rate in 2010.

Table 2-6—CAUTI per 1,000 Discharges (AHRQ)

2010 HAC Rates (per 1,000 discharges)	2011 HAC Rates (per 1,000 discharges)	2012 HAC Rates (per 1,000 discharges)	2013 HAC Rates (per 1,000 discharges)
12.21	11.30	10.58	8.85

Source: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html and Noel Eldridge, AHRQ Center for Quality Improvement and Safety, provided on August 28, 2015, for the final 2013 data.

CLABSI

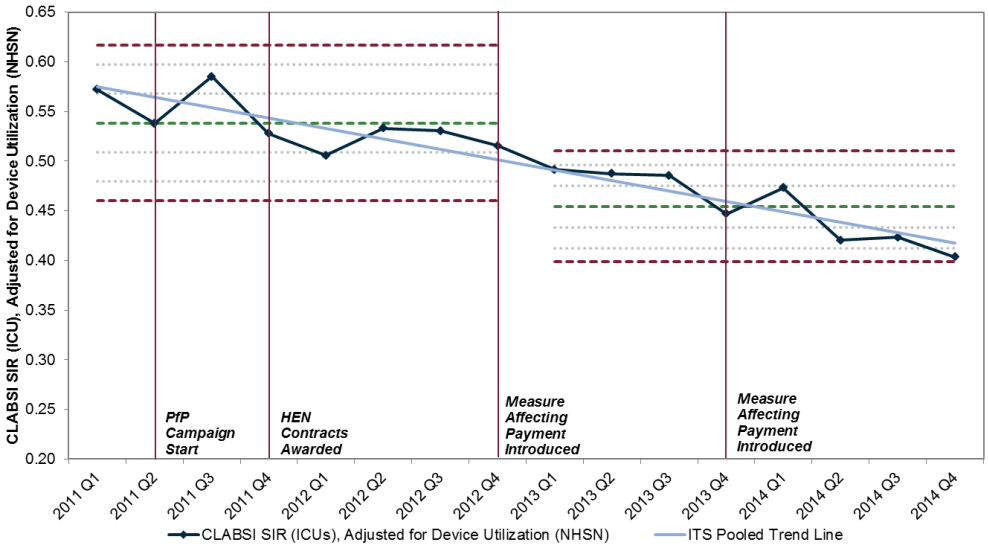
All-payer data from the AHRQ National Scorecard show 49.09 percent improvement since 2010 while less complete all-payer data from NDNQI indicate 24.46 percent improvement since Q1 2011, and the CLABSI SIR for ICU units from NHSN shows 29.50 percent improvement. Medicare claims-based data for CRBSI per 1,000 discharges exhibit a 56.08 percent decline between Q2 2011 and Q4 2014. The rate of CRBSI per 1,000 Medicaid adults declined by 17.40 percent between Q1 2009 and Q4 2010. The measure for the Medicaid pediatric population showed no signal of improvement from the control chart.[2-10]

Analysis of the speed with which the rate is changing shows no acceleration in the post-PfP period for any of the CLABSI measures presented below.

CLABSI SIR for ICUs, Adjusted for Device Utilization (NHSN) (Figure 2-22)

- Evidence suggesting improvement rate.
  - Shift in center line observed in Q1 2013.
  - Six consecutive reductions from Q2 2012 to Q4 2013.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.01 ($p < 0.05$).
- Rate decreased 29.50 percent between Q1 2011 and Q4 2014.

Figure 2-22—CLABSI SIR for ICUs, Adjusted for Device Utilization (NHSN)



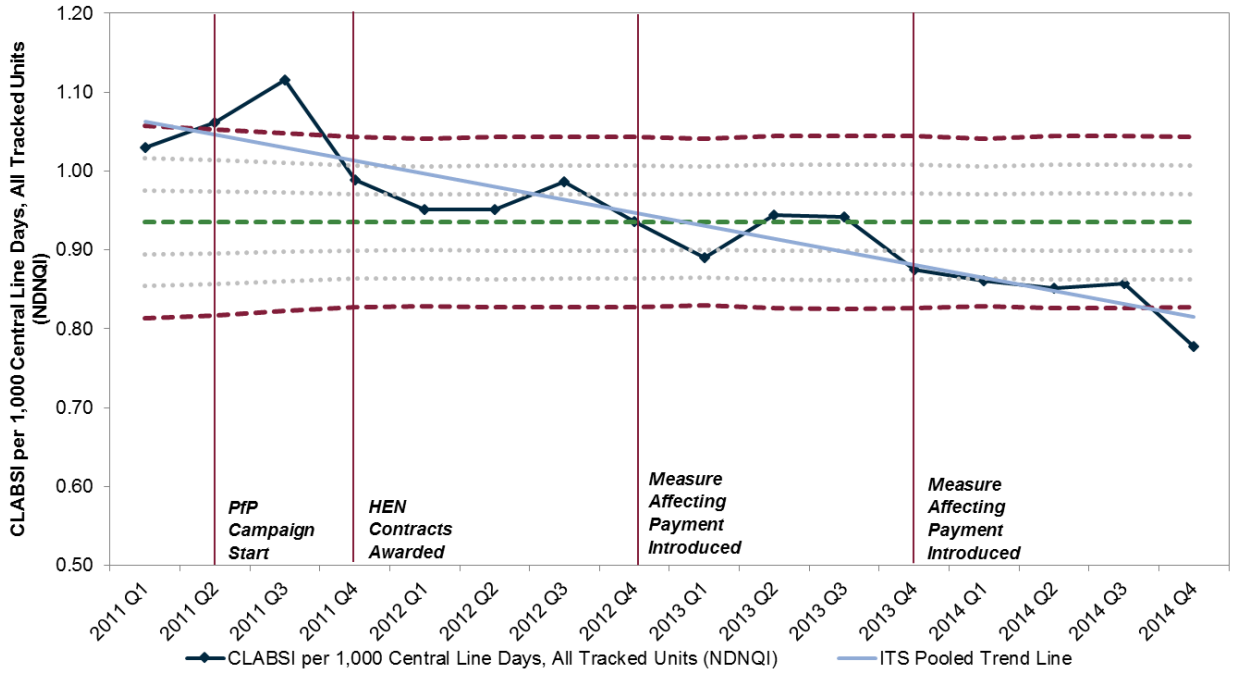
Source: NHSN. Data are between 3,048 and 3,152 hospitals per quarter. Data are adjusted for change in utilization, which also changes risk of harm. To adjust, the national unadjusted CLABSI SIR (ICUs) data points are multiplied by the ratio of the national device utilization rate for that period to the national device utilization rate in the first quarter.
Note: Center line and control limits (X chart) for the first phase were calculated with data between Q1 2011 and Q4 2012. Center line and control limits (X chart) for the second phase were calculated with data between Q1 2013 and Q4 2014. While the PfP campaign began in April 2011, HEN contracts were not awarded until December 2011; only modest educational support occurred in 2011. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. The data from NHSN were not provided with clustered standard errors that would account for the non-independence of discharges within the same facilities. Without the clustered standard errors, the SPC charts and ITS analysis performed are less conservative than would be the case if the correlation of outcomes within facilities were accounted for in the analysis.

2-10 CRBSIs are a subset of CLABSIs.

CLABSI per 1,000 Central Line Days, All Tracked Units (NDNQI) (Figure 2-23)

- Evidence suggesting special cause variation.
  - Q2 2011 and Q3 2011 fall above the upper control limit.
  - Last data point is below the lower control limit, suggesting special cause variation.
  - Data from Q1 2014 to Q3 2014 fall between the -2σ limit and the lower control limit.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.02 per 1,000 central line days ($p < 0.05$).
- Rate decreased by 24.46 percent between Q1 2011 and Q4 2014.
- IPPS payment incentive related to Vascular Catheter-Associated Infections became effective in October 2012.
- CLABSI incorporated into Hospital Value-Based Purchasing (VBP) Program in October 2013.

Figure 2-23—CLABSI per 1,000 Central Line Days (NDNQI)

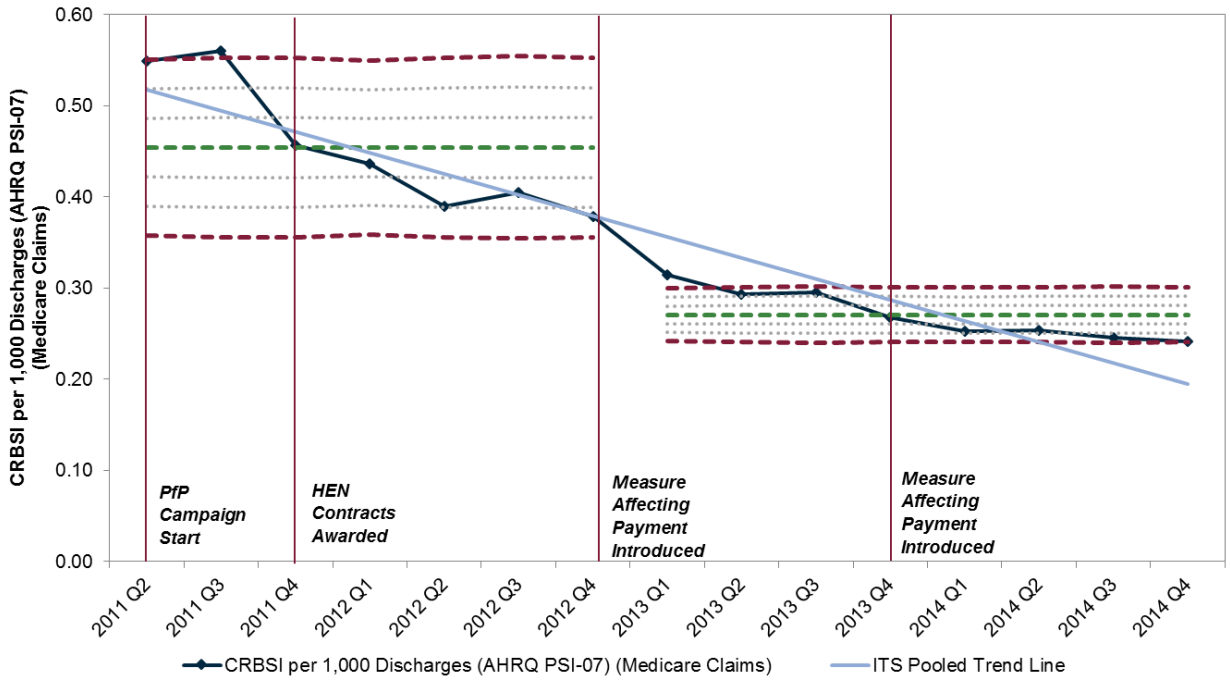


Source: NDNQI. Data are from between 635 and 745 hospitals per quarter.
Note: Control limits and center line (U′ chart) for period Q1 2011 to Q4 2014 were constructed using data from Q1 2011 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

CRBSI per 1,000 Discharges (AHRQ PSI-07) (Medicare Claims) (Figure 2-24)

- Evidence suggesting improvement in rate.
  - Shift in center line observed in Q1 2013.
  - Data for Q3 2011 fall above the first phase upper control limit.
  - Data for Q1 2013 fall above the second phase upper control limit.
  - Data for Q2 2013 and Q3 2013 fall between the second phase 2σ limit and the upper control limit.
  - Data for Q3 2014 to Q4 2014 fall between the second phase -2σ limit and the lower control limit.
- Evidence that the measure trended down without a significant break.
  - The average quarterly change was -0.02 per 1,000 discharges ($p < 0.05$).
- Rate decreased 56.08 percent between Q2 2011 and Q4 2014.
- IPPS payment incentive related to Vascular Catheter-Associated Infections became effective in October 2012.

Figure 2-24—CRBSI per 1,000 Discharges (Medicare Claims)



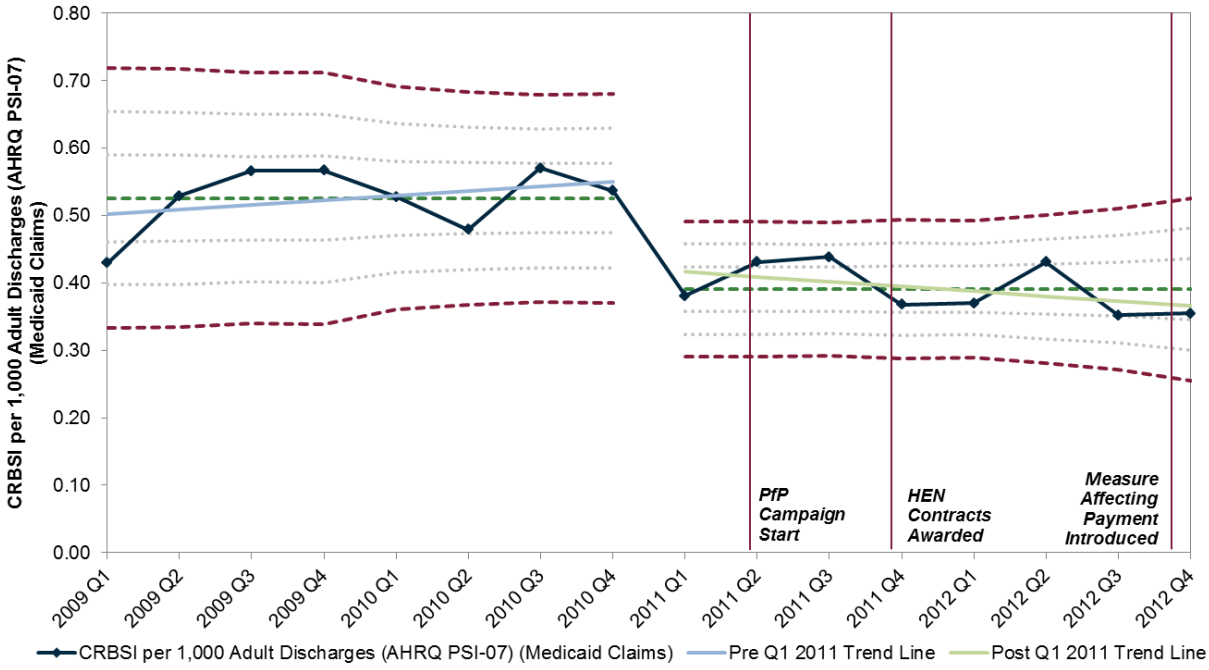
Source: Rates calculated by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS based on Medicare FFS claims data. The Evaluation Contractor also conducted a similar analysis, adding a longer time series (from 2009) and using regression analysis to control for changing demographics over time. The data showed a similar pattern.

Note: Control limits and center line (U′ chart) for the first phase were calculated with data between Q2 2011 to Q4 2012. Control limits and center line (U' chart) for the second phase were calculated with data between Q1 2013 to Q4 2014. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma lines; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicare FFS claims for all hospitals that reliably report POA status (≥ 95 percent of the hospital's diagnoses for a given quarter are accompanied by a valid code for POA) and that have the following characteristics: all hospitals paid under Medicare's IPPS, CAH, cancer hospitals, and Maryland hospitals. Data include between 1,444,306 and 1,633,261 discharges per quarter.

CRBSI per 1,000 Adult Discharges (AHRQ PSI-07) (Medicaid Claims) (Figure 2-25)

- Evidence suggesting improvement in rate.
  - Shift in center line observed in Q1 2011.
- Evidence that the measure trended down without a significant break.
  - The average rate was shifted 0.13 ($p < 0.05$) lower in Q1 2011.
- Rate decreased 32.88 percent between CY 2010 and Q4 2012.
- IPPS payment incentive related to Vascular Catheter-Associated Infections became effective in October 2012.

Figure 2-25—CRBSI per 1,000 Adult Discharges (Medicaid Claims)

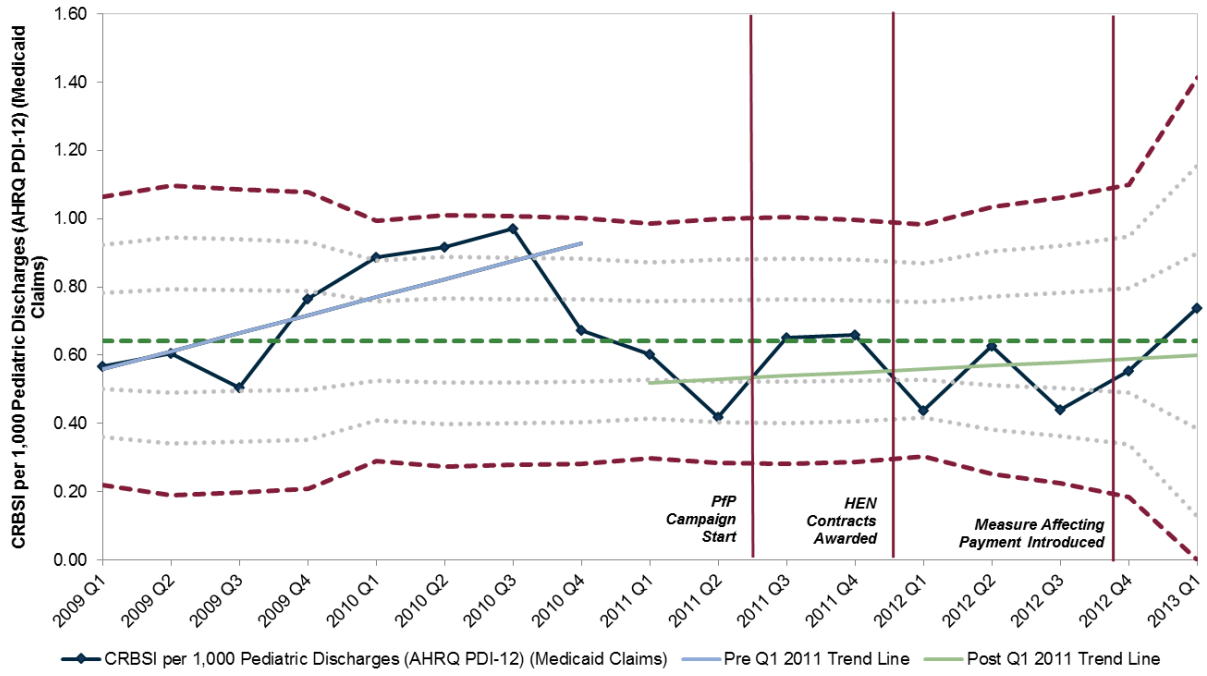


Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U chart) for the period Q1 2009 to Q4 2010 constructed using data from Q1 2009 through Q4 2010. Control limits and center line (U' chart) for the period Q1 2011 to Q4 2012 constructed using data from Q1 2011 through Q4 2012. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states), assuming all diagnosis codes were not POA (Medicaid claims do not include POA data) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 107,971 and 202,016 discharges per quarter.

CRBSI per 1,000 Pediatric Discharges (AHRQ PDI-12) (Medicaid Claims) (Figure 2-26)

- Evidence suggesting special cause variation.
  - Q1 to Q3 2010 fall between the 2σ limit and upper control limit.
- Evidence of a significant break in the rate.
  - The average rate was shifted 0.42 ($p < 0.01$) lower in Q1 2011.
- Rate increased 14.40 percent between CY 2010 and Q1 2013.

Figure 2-26—CRBSI per 1,000 Pediatric Discharges (Medicaid Claims)



Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q1 2013. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states), assuming all diagnosis codes were not POA (Medicaid claims do not include POA data) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 15,111 and 77,715 discharges per quarter. Medicaid data included in these analyses comprise claims for nondual eligible patients ages zero to 17.

CLABSI per 1,000 Discharges (AHRQ National Scorecard) (Table 2-7)

- Evidence for improvement from raw rates in this measure.
  - Steady but modest decline in rates between 2010 and 2011 and/or 2012 (7.27 percent improvement between 2010 and 2012). Acceleration in the decline during 2013 to 0.28 per 1,000 discharges compared to 0.55 in 2010 (49.09 percent improvement), and 45.10 percent improvement since 2012.

Table 2-7—CLABSI per 1,000 Discharges (AHRQ)

2010 HAC Rates (per 1,000 discharges)	2011 HAC Rates (per 1,000 discharges)	2012 HAC Rates (per 1,000 discharges)	2013 HAC Rates (per 1,000 discharges)
0.55	0.52	0.51	0.28

Source: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html and Noel Eldridge, AHRQ Center for Quality Improvement and Safety, provided on August 28, 2015, for the final 2013 data.

Falls

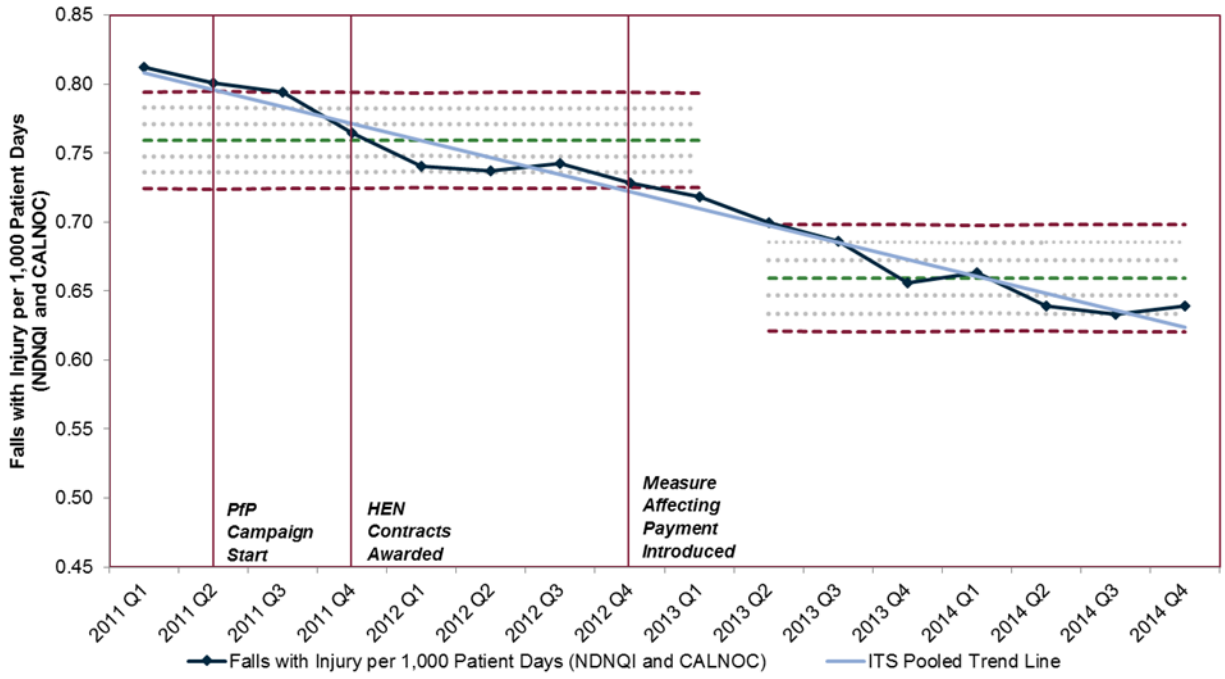
The two all-payer measures from NDNQI and CALNOC data, which are more sensitive than claims-based measures but which include fewer hospitals, show evidence of improvement—21.32 percent for all falls and 9.46 percent for falls with injury between Q1 2011 and Q4 2014. The AHRQ National Scorecard measure also shows improvement in all falls—9.32 percent since 2010. For the Medicare population, the narrow post-operative hip fracture measure rate per 1,000 discharges (AHRQ PSI-08) shows no evidence of change.

Analysis of the speed with which the rate is changing finds no change in the post-PfP period relative to the baseline period.

Falls with Injury per 1,000 Patient Days (NDNQI and CALNOC) (Figure 2-27)

- Evidence suggesting improvement in rate.
  - Q1 and Q2 2011 fall above the upper control limit.
  - Shift in center line observed in Q2 2013.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change in rate was -0.01 per 1,000 patient days ($p < 0.05$).
- Rate decreased 21.32 percent between Q1 2011 and Q4 2014.
- IPPS payment incentive related to fall trauma became effective in October 2012.

Figure 2-27—Falls with Injury per 1,000 Patient Days (NDNQI and CALNOC)

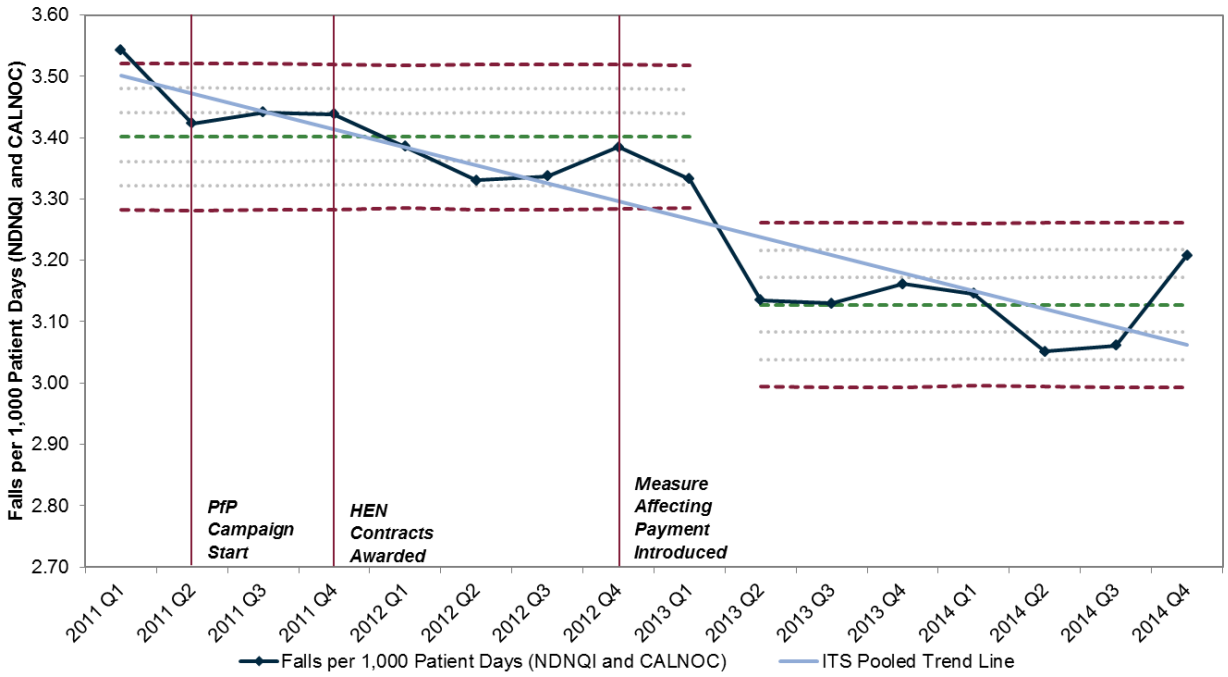


Source: NDNQI and CALNOC. Data are between 1,268 and 1,342 hospitals per quarter.
Note: Control limits and center line (U′ chart) for period Q1 2011 to Q1 2013 were constructed using data from Q1 2011 to Q1 2013. Control limits and center line (U′ chart) for period Q2 2013 to Q4 2014 were constructed using data from Q2 2013 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Falls per 1,000 Patient Days (NDNQI and CALNOC) (Figure 2-28)

- Evidence suggesting improvement in rate.
  - Q1 2011 falls above the upper control limit.
  - Shift in center line observed in Q2 2013.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.03 per 1,000 patient days ($p < 0.05$).
- Rate decreased by 9.46 percent between Q1 2011 and Q4 2014.
- IPPS payment incentive related to fall trauma became effective in October 2012.

Figure 2-28—Falls per 1,000 Patient Days (NDNQI and CALNOC)

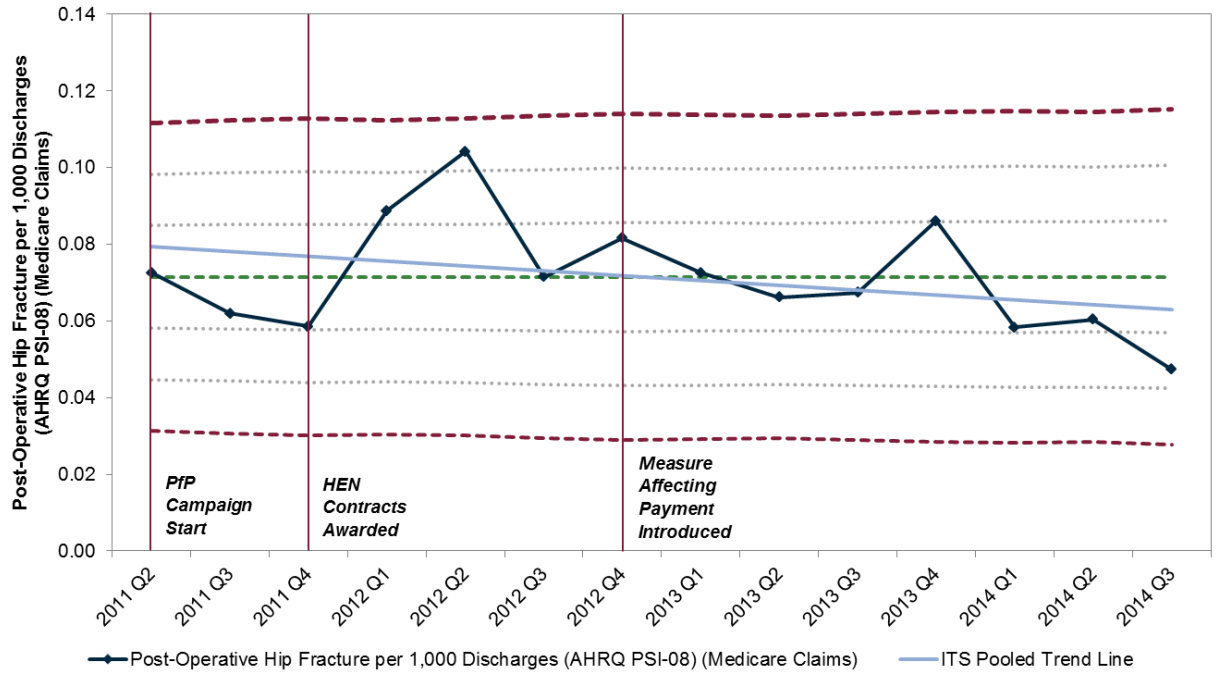


Source: NDNQI and CALNOC. Data are between 1,268 and 1,342 hospitals per quarter.
Note: Control limits and center line (U′ chart) for period Q1 2011 to Q1 2013 were constructed using data from Q1 2011 to Q1 2013. Control limits and center line (U′ chart) for period Q2 2013 to Q4 2014 were constructed using data from Q2 2013 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Post-Operative Hip Fracture per 1,000 Discharges (AHRQ PSI-08) (Medicare Claims) (Figure 2-29)

- No evidence for special cause variation.
  - Control limits are temporary pending additional data.[2-11]
- No evidence of significant break or trend in the rate.
- Rate decreased 34.75 percent between Q2 2011 and Q3 2014.
- IPPS payment incentive related to fall trauma became effective in October 2012.

Figure 2-29—Post-Operative Hip Fracture per 1,000 Discharges (Medicare Claims)



Source: Rates calculated by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS based on Medicare FFS claim data. The Evaluation Contractor processed and analyzed the claims data using a longer time series (from 2009) and adjusting for changing demographics over time. The analysis showed a similar pattern.
Note: Control limits and center line (U chart) were constructed using data from Q2 2011 to Q3 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two sigma limits. Calculations are based on Medicare FFS claims for all hospitals that reliably report POA status (≥ 95 percent of the hospital's diagnoses for a given quarter are accompanied by a valid code for POA) and that have the following characteristics: all hospitals paid under Medicare's IPPS, CAH, cancer hospitals, and Maryland hospitals. Data include between 337,812 and 399,520 surgical discharges per quarter.

[2-11] For SPC charts with fewer than 16 observations, the control limits are unstable, and thus final control limits cannot be calculated without additional data.

Falls per 1,000 Discharges (AHRQ National Scorecard) (Table 2-8)

- Evidence for improvement from the raw rates in this measure.
  - Small decline (1.76 percent improvement) in rates from 2010 to 2011; sharper decline in 2012 and 2013 (9.32 percent improvement compared to 2010).

Table 2-8—Falls per 1,000 Discharges (AHRQ)

2010 HAC Rates (per 1,000 discharges)	2011 HAC Rates (per 1,000 discharges)	2012 HAC Rates (per 1,000 discharges)	2013 HAC Rates (per 1,000 discharges)
7.94	7.80	7.16	7.20

Source: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html and Noel Eldridge, AHRQ Center for Quality Improvement and Safety, provided on August 28, 2015, for the final 2013 data.

OB-Other

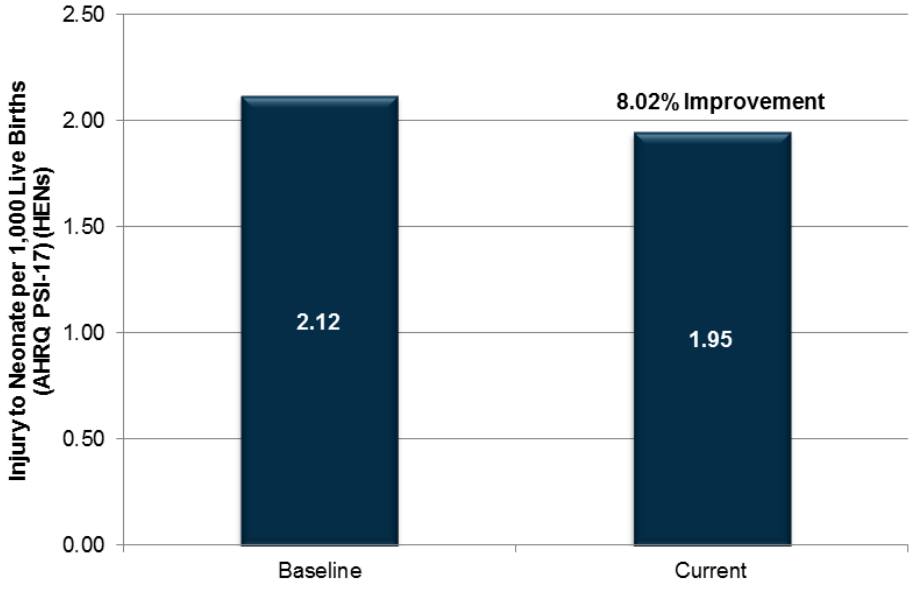
OB harm measures available for consideration were OB trauma (as measured by AHRQ PSI-18 and PSI-19), and injury to neonate (as measured by AHRQ PSI-17). OB trauma during vaginal delivery both with and without instrument (AHRQ PSI-18 and PSI-19) showed improvement in rates over time in HEN-reported data (9.49 percent and 17.38 percent, respectively) when examined with *z*-tests for differences in proportions. However, the Medicaid data through 2012 did not exhibit a significant downward trend for PSI-18. In contrast, there was a downward turn in trend for PSI-19 in Q1 2012. The combined OB trauma data from the AHRQ National Scorecard suggest improvement of 5.60 percent between 2010 and 2012. HENs' all-payer data on injury to neonate show no significant improvement, while the trend in the Medicaid population data shows a significant average quarterly downward trend.

Analysis of the speed with which the rate is changing finds no change in the post-PfP period relative to the baseline period for two of the three Medicaid measures: injury to neonate (PSI-17), or OB trauma (PSI-18).

Injury to Neonate per 1,000 Live Births (AHRQ PSI-17) (HENs) (Figure 2-30)

- No evidence for significant improvement.
  - While the most recent or current estimated rate is slightly lower than the baseline rate (by 8.02 percent or 0.17 per 1,000 live births), this difference was not statistically significant.

Figure 2-30—Injury to Neonate per 1,000 Live Births (HENs)

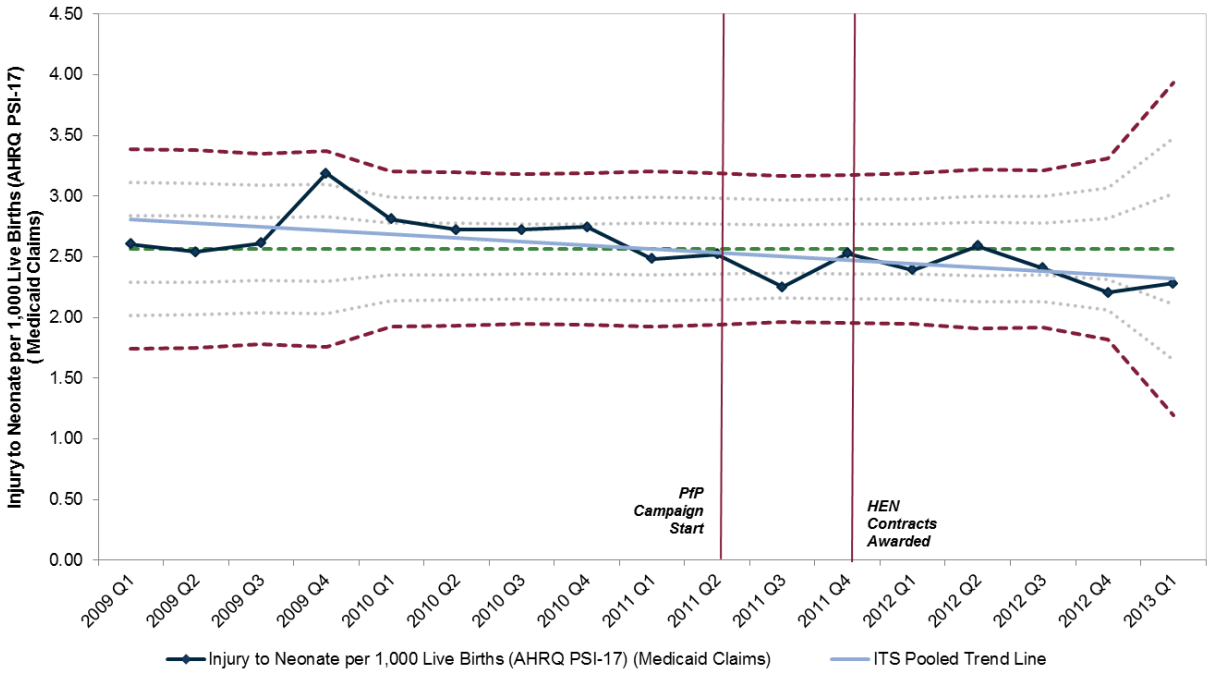


Source: HEN-reported data submitted November 2014.
Note: N = 1,466 hospitals in 19 HENs. Among the 19 HENs reporting on this measure (one of which has two participating cohorts), 12 baseline periods are in 2010, five began in 2011, and three began in 2012. Nine of the baseline periods are based on quarterly data, two are based on 6 months of data, and nine are based on annual data. Fourteen current periods end in Q1 2014 or later, five end in 2013, and one ends in 2010. One current period is 6 months and the rest are 3 months.

Injury to Neonate per 1,000 Live Births (AHRQ PSI-17) (Medicaid Claims) (Figure 2-31)

- No evidence to suggest special cause variation.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.03 per 1,000 live births ($p < 0.05$).
- Rate decreased 17.11 percent between CY 2010 and Q1 2013.

Figure 2-31—Injury to Neonate per 1,000 Live Births (Medicaid Claims)

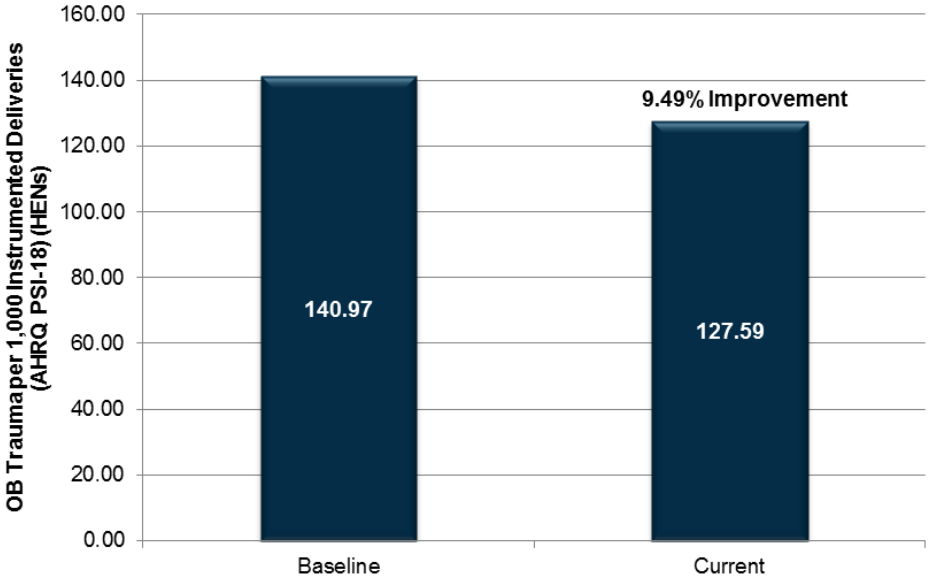


Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U chart) constructed using data from Q1 2009 to Q1 2013. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 12,248 and 63,122 discharges per quarter.

Obstetric Trauma, Vaginal Delivery with Instrument, per 1,000 Instrumented Deliveries (AHRQ PSI-18) (HENs) (Figure 2-32)

- Evidence for significant improvement.
  - The most recent or current rate is 9.49 percent lower (or 13.38 per 1,000 instrumented deliveries) than the baseline rate.

Figure 2-32—Obstetric Trauma per 1,000 Instrumented Deliveries (HENs)

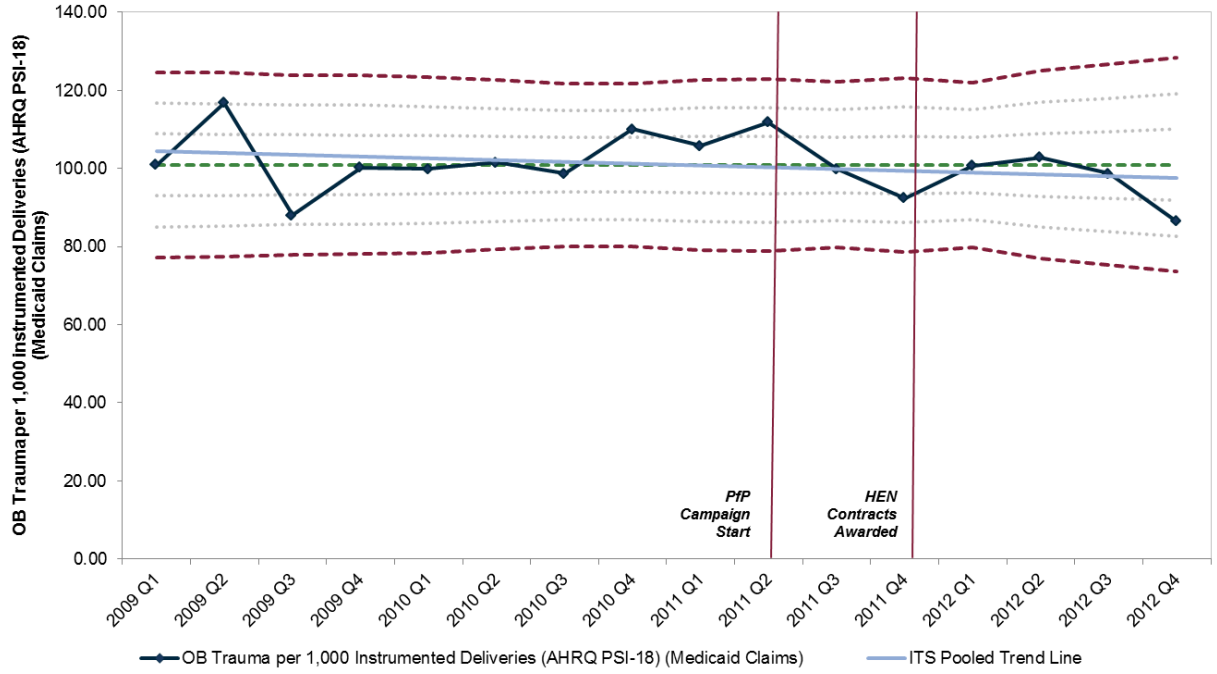


Source: HEN-reported data submitted November 2014.
Note: N = 1,633 hospitals in 23 HENs. Among the 23 HENs reporting on this measure (one of which has two participating cohorts), 16 baseline periods begin in 2010, five begin in 2011, and two begin in 2012. Twelve baseline periods are based annual data, seven are based on 3 months of data, two are based on 6 months of data, and two are based on more than a year of data. Nineteen current periods end in Q1 2014 or later, three end in 2013, and one ends in Q4 2012; 21 current periods are 3 months, one is less than a year, and one is more than a year.

Obstetric Trauma, Vaginal Delivery with Instrument, per 1,000 Instrumented Deliveries (AHRQ PSI-18) (Medicaid Claims) (Figure 2-33)

- No evidence to suggest special cause variation.
- No evidence of significant break in level or trend in the rate.
- Rate decreased 15.74 percent between CY 2010 and Q4 2012.

Figure 2-33—Obstetric Trauma per 1,000 Instrumented Deliveries (Medicaid Claims)

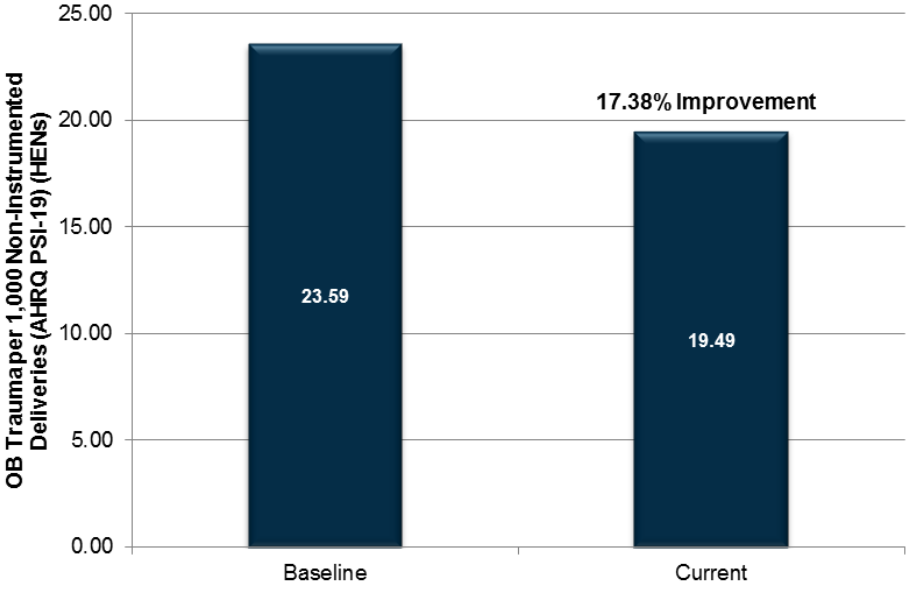


Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q4 2012. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 1,409 and 2,416 discharges per quarter.

Obstetric Trauma, Vaginal Delivery without Instrument, per 1,000 Non-Instrumented Deliveries (AHRQ PSI-19) (HENs) (Figure 2-34)

- Evidence for significant improvement.
  - The most recent or current rate is 17.38 percent (or 4.10 per 1,000 non-instrumented deliveries) lower than the baseline rate.

Figure 2-34—Obstetric Trauma per 1,000 Non-Instrumented Deliveries (HENs)

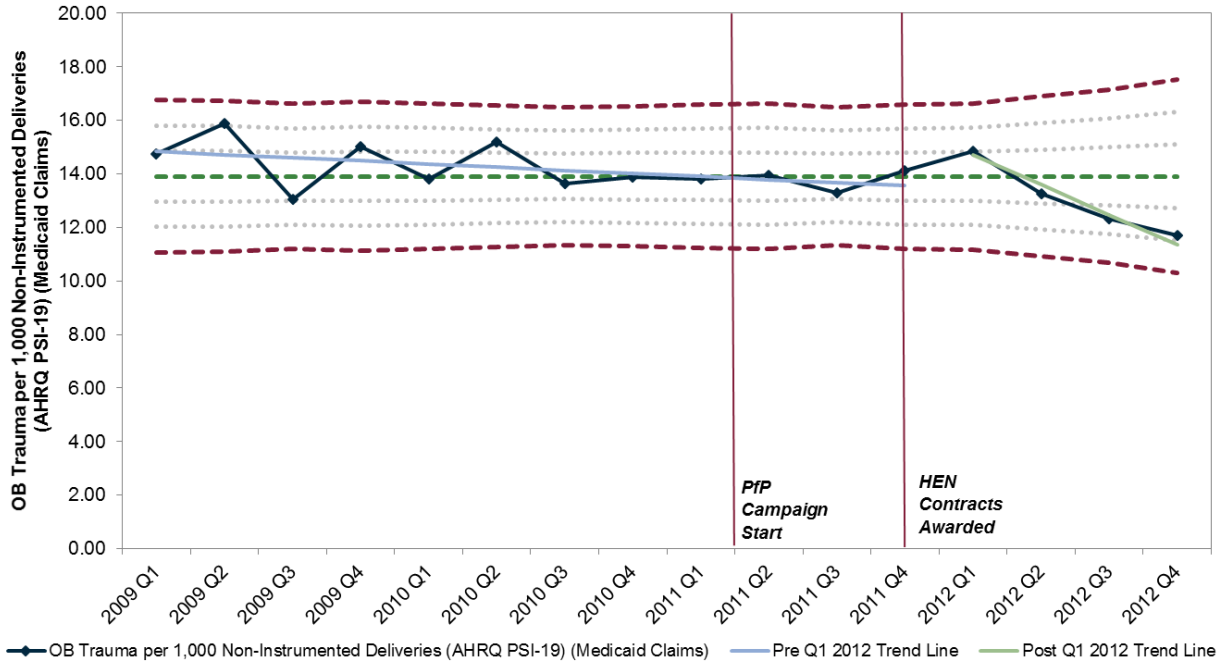


Source: HEN-reported data submitted November 2014.
Note: N = 1,759 hospitals in 22 HENs. Of the 22 HENs, 16 HEN baseline periods begin in 2010; five begin in 2011, and one begins in 2012. Most of the baseline periods are based on annual data, eight HENs use less than a full year but at least one quarter of data, and two HENs use more than a year of data. Eighteen HENs' current periods end in Q1 2014 or later and four HENs end in 2013; most current periods are 3 months, one HEN is 6 months, and one HEN is more than a year.

Obstetric Trauma, Vaginal Delivery without Instrument, per 1,000 Non-Instrumented Deliveries (AHRQ PSI-19) (Medicaid Claims) (Figure 2-35)

- No evidence to suggest special cause variation.
- Evidence that the level has worsened after the HEN contracts were awarded, but this was offset by a trend that significantly improved after Q1 2012.
  - The average quarterly linear change was -1.00 percent greater ($p < 0.01$) after Q1 2012 than before.
- Rate decreased 17.27 percent between CY 2010 and Q4 2012.

Figure 2-35—Obstetric Trauma per 1,000 Non-Instrumented Deliveries (Medicaid Claims)

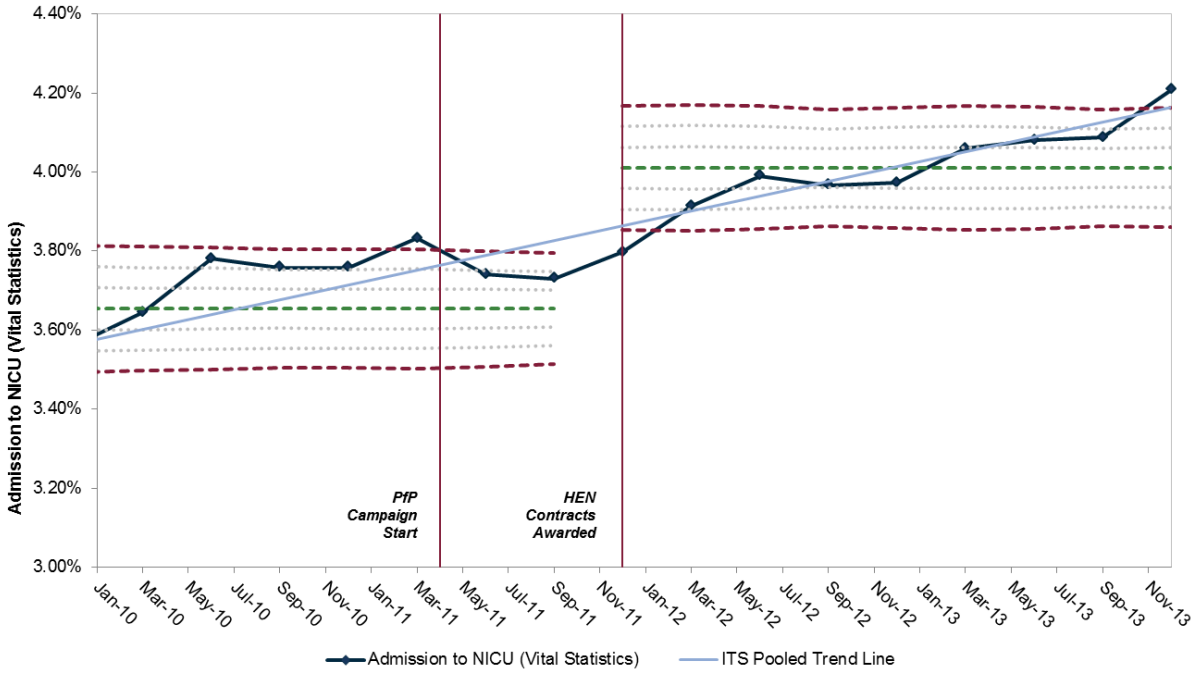


Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q4 2012. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 21,410 and 42,577 discharges per quarter.

Admission to Neonatal Intensive Care Unit (NICU) (Vital Statistics) (Figure 2-36)

- Evidence suggesting worsening in rate.
  - Shift in center line observed in Q4 2011.
- The average quarterly linear change was 0.04 percent ($p < 0.001$).
- Rate increased by 12.61 percent between CY 2010 and Q4 2013.

Figure 2-36—Admission to NICU (Vital Statistics)



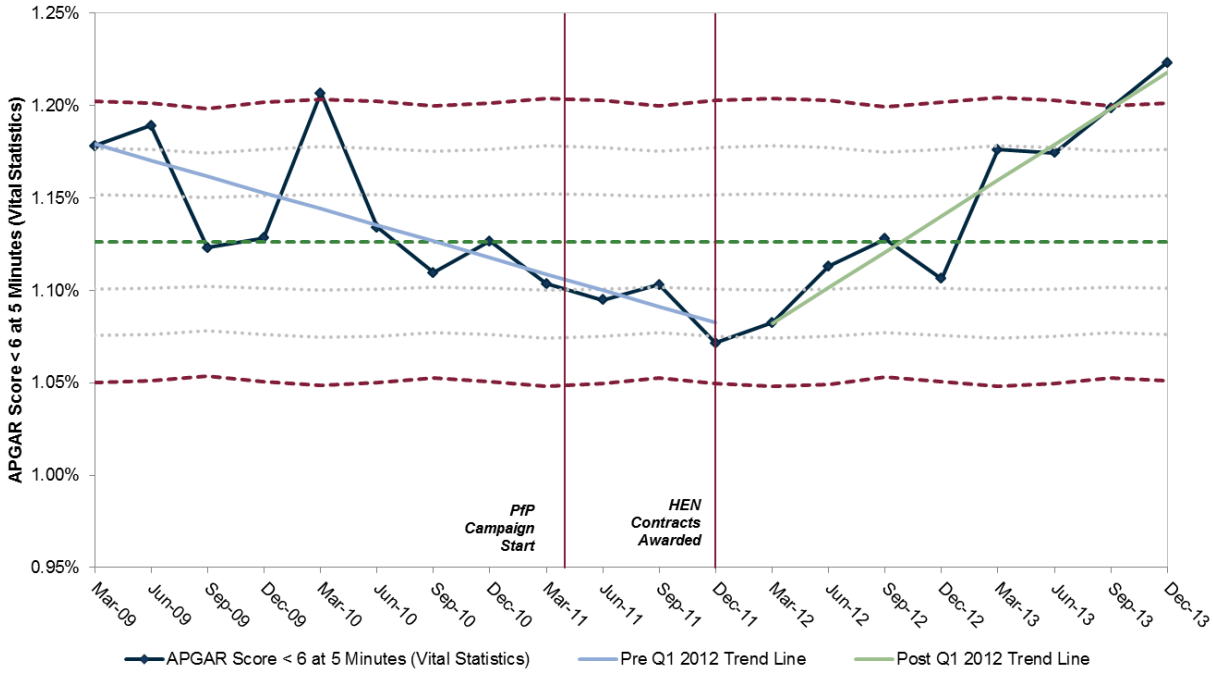
Source: NVSS.

Note: Center line and control limits (U' chart) for first phase were calculated with data between Q1 2009 and Q3 2011. Center line and control limits (U' chart) for second phase were calculated with data between Q4 2011 and Q4 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

APGAR Score < 6 at 5 Minutes (Vital Statistics) (Figure 2-37)

- Evidence of special cause variation.
  - Data for Q1 2010 and Q4 2013 fall above the upper control limit.
- Evidence that the rate has worsened since Q1 2012.
  - The average quarterly change was a 0.03 percent ($p < 0.001$) greater increase after Q1 2012 than before.
- Rate increased by 7.01 percent between CY 2010 and Q4 2013.

Figure 2-37—APGAR Score < 6 at 5 Minutes (Vital Statistics)



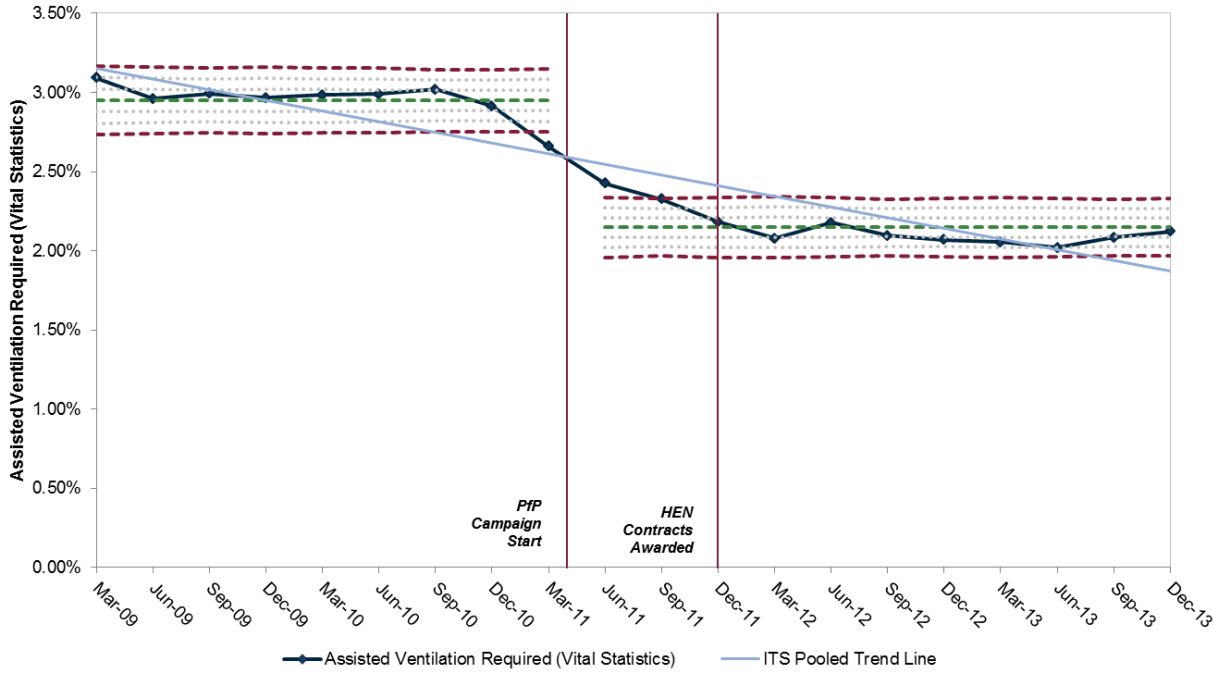
Source: NVSS.

Note: Center line and control limits (U' chart) were calculated with data between Q1 2009 and Q3 2012. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Assisted Ventilation Required (Vital Statistics) (Figure 2-38)

- Evidence suggesting improvement in rate.
  - Shift in center line observed in Q2 2011.
- The average quarterly linear change was -0.06 percent ($p < 0.001$).
- Rate decreased by 28.72 percent between CY 2010 and Q4 2013.

Figure 2-38—Assisted Ventilation Required (Vital Statistics)



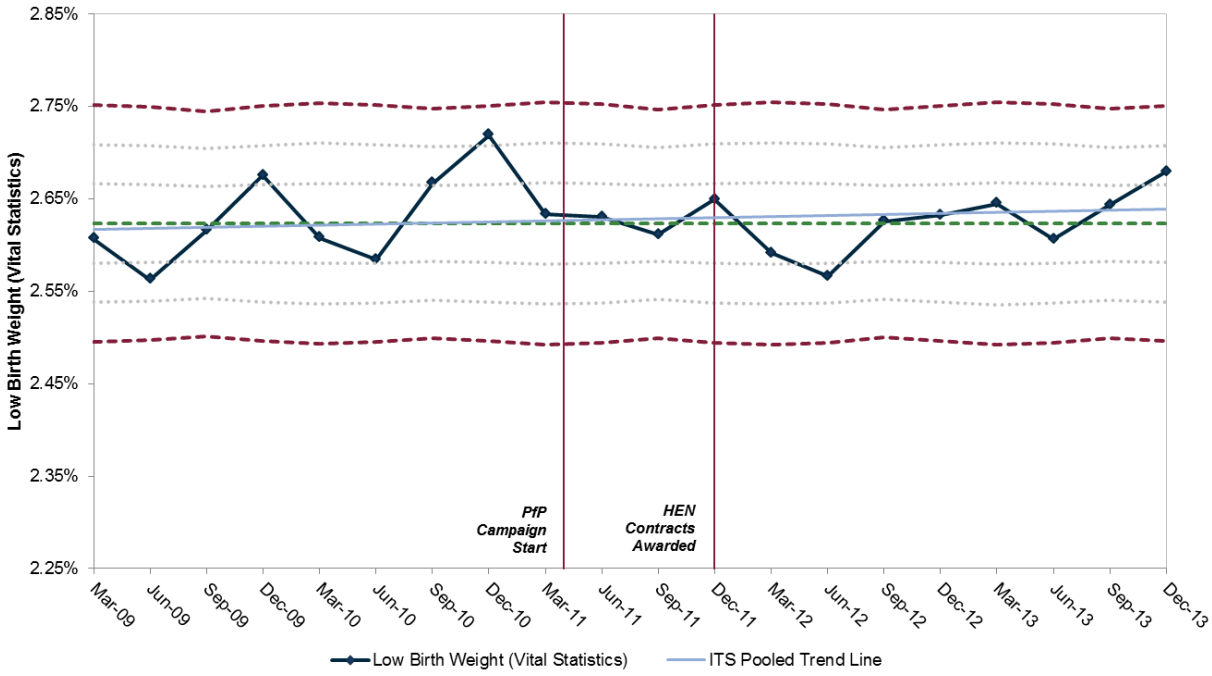
Source: NVSS.

Note: Center line and control limits (U' chart) for first phase were calculated with data between Q1 2009 and Q1 2011. Center line and control limits (U' chart) for second phase were calculated with data between Q2 2011 and Q4 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Low Birth Weight (Vital Statistics) (Figure 2-39)

- No evidence of special cause variation.
- No evidence of significant break or trend in the rate.
- Rate increased by 1.28 percent between CY 2010 and Q4 2013.

Figure 2-39—Low Birth Weight (Vital Statistics)



Source: NVSS.
Notes: Center line and control limits (U' chart) were calculated with data between Q1 2009 and Q3 2012. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Obstetric Trauma per 1,000 Discharges (AHRQ PSI-18 and PSI-19) (AHRQ National Scorecard) (Table 2-9)[2-12]

- Evidence for improvement from the raw rates in this measure.
  - Small decline (5.60 percent) in rates from 2010 to 2013.

Table 2-9—OB Trauma per 1,000 Discharges (AHRQ)

2010 HAC Rates (per 1,000 discharges)	2011 HAC Rates (per 1,000 discharges)	2012 HAC Rates (per 1,000 discharges)	2013 HAC Rates (per 1,000 discharges)
2.50	2.51	2.36	2.36

Source: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html and Noel Eldridge, AHRQ Center for Quality Improvement and Safety, provided on August 28, 2015, for the final 2013 data.

2-12 The use of "per 1,000 discharges" as the denominator could result in a misleading trend if the birth rate is unstable over the years trended. However, the birth rate has been stable during 2010-2012 (64.1 per 1,000 women ages 15-44 in 2010, 63.2 in 2011, and 63.0 in 2012). Available at http://www.childtrends.org/databank/indicators-by-topic-area/demographics/

Pressure Ulcers

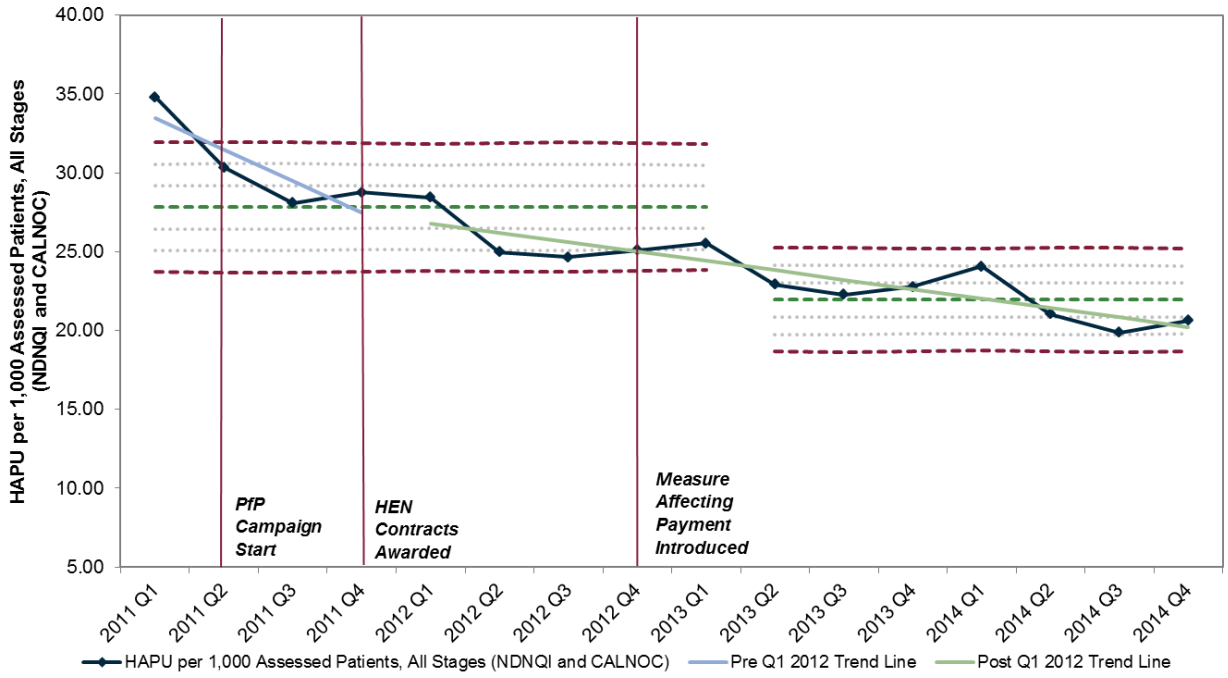
The broader measures of pressure ulcers in the available all-payer data suggest substantial progress; the AHRQ National Scorecard measure for all pressure ulcers showed 19.35 percent improvement—about half of the decline observed in the less complete NDNQI and CALNOC data. However, progress is not found or is less certain for measures of the most severe pressure ulcers (stages 3+). The all-payer measure's graph shows a significant linear decline since Q2 2011, but no clear structural break in the series. The Medicare claims-based measure showed improvement in the average linear trend, but no structural breaks were identified and the control chart does not show evidence for improvement. Although the rate of severe pressure ulcers for the Medicaid population may have increased relative to baseline, the Medicaid data do not contain POA indicators, so it is uncertain whether the increase stemmed from an increase in pressure ulcers acquired in the hospital.

Analyses of the speed with which the rate is changing finds mixed results depending on the measure and population—the rate of improvement slowed for the all-payer all-stage NDNQI trends in the post-PfP period, no change in the speed of change for the all-payer NDNQI trend in the most severe pressure ulcers, and the Medicare and Medicaid trends of the most severe pressure ulcers.

Hospital-Acquired Pressure Ulcer per 1,000 Assessed Patients, All Stages (NDNQI and CALNOC) (Figure 2-40)

- Evidence suggesting improvement in rate.
  - Q1 2011 falls above the upper control limit.
  - Q2 and Q3 2012 data fall between -2σ and the lower control limit.
  - Shift in center line observed in Q2 2013.
- Evidence that the trend has worsened since Q1 2012.
  - The average quarterly rate change was a 1.41 per 1,000 assessed patients ($p < 0.05$) smaller decline after Q1 2012 than before.
- Rate decreased by 40.72 percent between Q1 2011 and Q4 2014.
- IPPS payment incentive related to pressure ulcers stages 3 and 4 became effective in October 2012.

Figure 2-40—Hospital-Acquired Pressure Ulcer per 1,000 Assessed Patients (All Stages) (NDNQI and CALNOC)

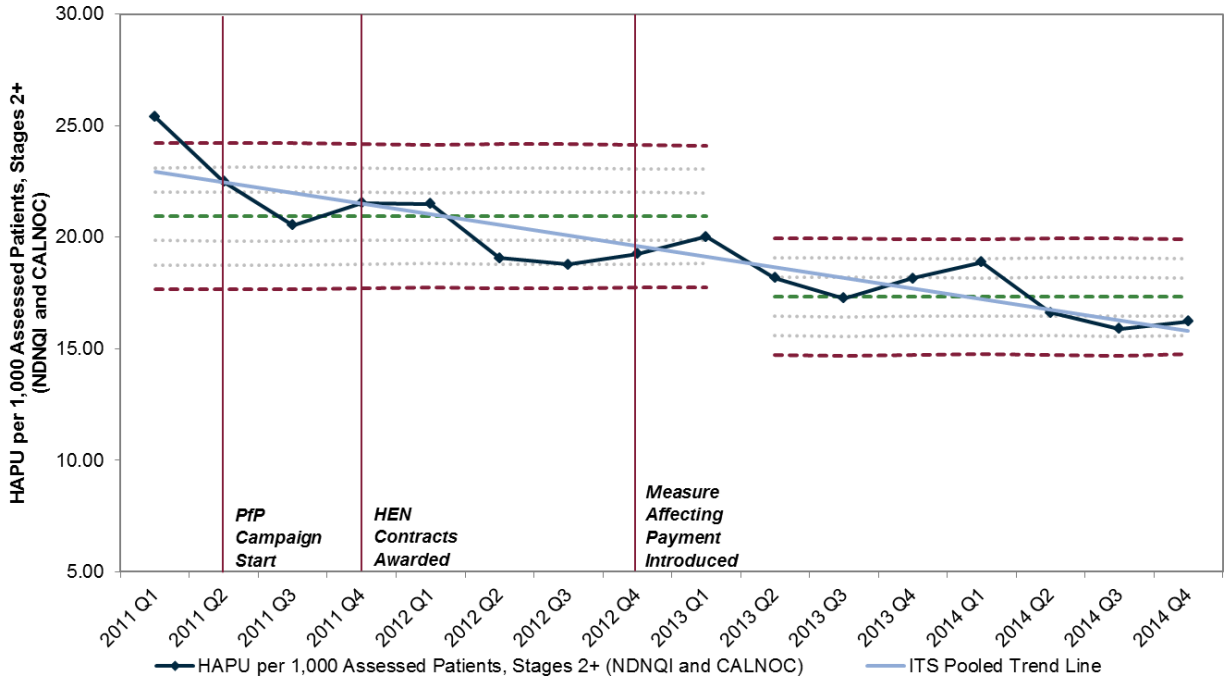


Source: NDNQI and CALNOC. Data are from between 1,291 and 1,357 hospitals per quarter.
Note: Control limits and center line (U′ chart) for period Q1 2011 to Q1 2013 were constructed using data from Q1 2011 to Q1 2013. Control limits and center line (U′ chart) for period Q2 2013 to Q4 2014 were constructed using data from Q2 2013 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Hospital-Acquired Pressure Ulcer per 1,000 Assessed Patients, Stages 2+ (NDNQI and CALNOC) (Figure 2-41)

- Evidence suggesting improvement in rate.
  - Q1 2011 falls above the upper control limit.
  - Shift in center line observed in Q2 2013.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.48 per 1,000 assessed patients ($p < 0.05$).
- Rate decreased by 36.10 percent between Q1 2011 and Q4 2014.
- IPPS payment incentive related to pressure ulcers stages 3 and 4 became effective in October 2012.

Figure 2-41—Hospital-Acquired Pressure Ulcer (Stages 2+) per 1,000 Assessed Patients (NDNQI and CALNOC)



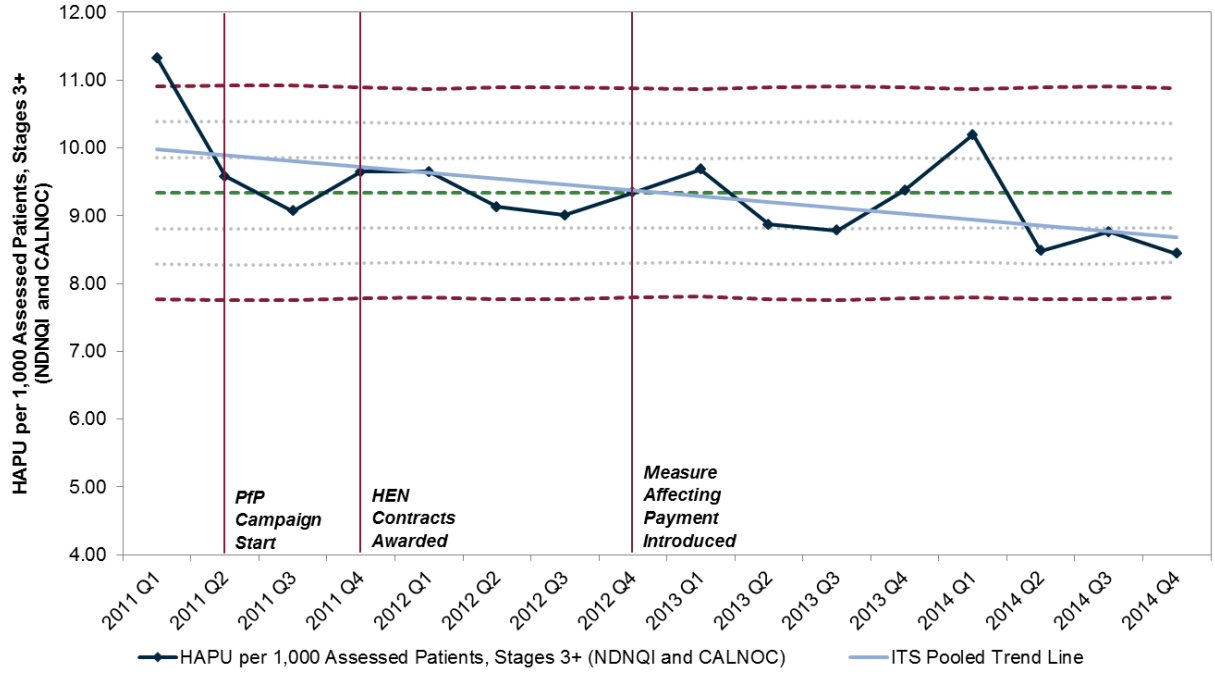
Source: NDNQI and CALNOC. Data are from between 1,291 and 1,357 hospitals per quarter.

Note: Control limits and center line (U′ chart) for period Q1 2011 to Q1 2013 were constructed using data from Q1 2011 to Q1 2013. Control limits and center line (U′ chart) for period Q2 2013 to Q4 2014 were constructed using data from Q2 2013 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Hospital-Acquired Pressure Ulcer per 1,000 Assessed Patients, Stages 3+ (NDNQI and CALNOC) (Figure 2-42)

- No recent evidence of special cause variation.
  - Q1 2011 falls above the upper control limit.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.09 per 1,000 assessed patients ($p < 0.05$).
- Rate decreased by 25.45 percent between Q1 2011 and Q4 2014.
- IPPS payment incentive related to pressure ulcers stages 3 and 4 became effective in October 2012.

Figure 2-42—Hospital-Acquired Pressure Ulcer (Stages 3+) per 1,000 Assessed Patients (NDNQI and CALNOC)

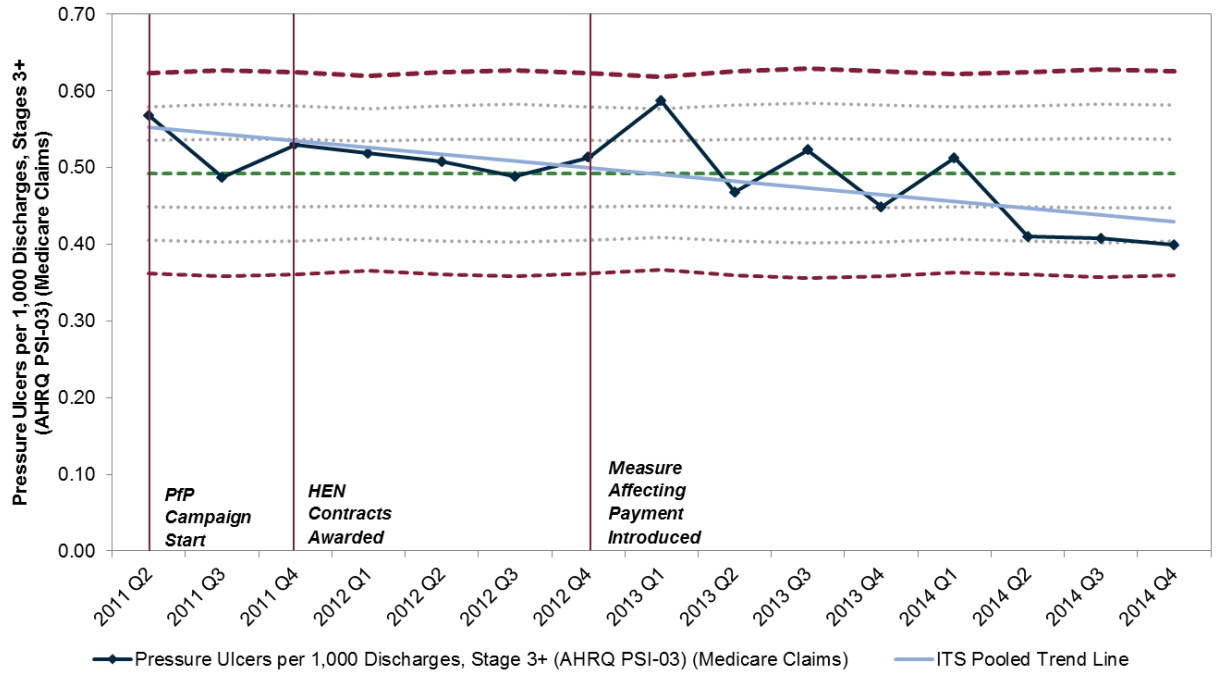


Source: NDNQI and CALNOC. Data are from between 1,291 and 1,357 hospitals per quarter.
Note: Control limits and center line (U′ chart) constructed using data from Q1 2011 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Pressure Ulcers per 1,000 Discharges, Stages 3+ (AHRQ PSI-03), Medicare Claims (Figure 2-43)

- No evidence of special cause variation.
- Evidence that the measure trended down without a significant break.
  - The average quarterly linear change was -0.01 per 1,000 discharges ($p < 0.05$).
- Rate decreased 29.80 percent between Q2 2011 and Q4 2014.
- IPPS payment incentive related to pressure ulcers stages 3 and 4 became effective in October 2012.

Figure 2-43—Pressure Ulcer (Stages 3+) per 1,000 Discharges (Medicare Claims)



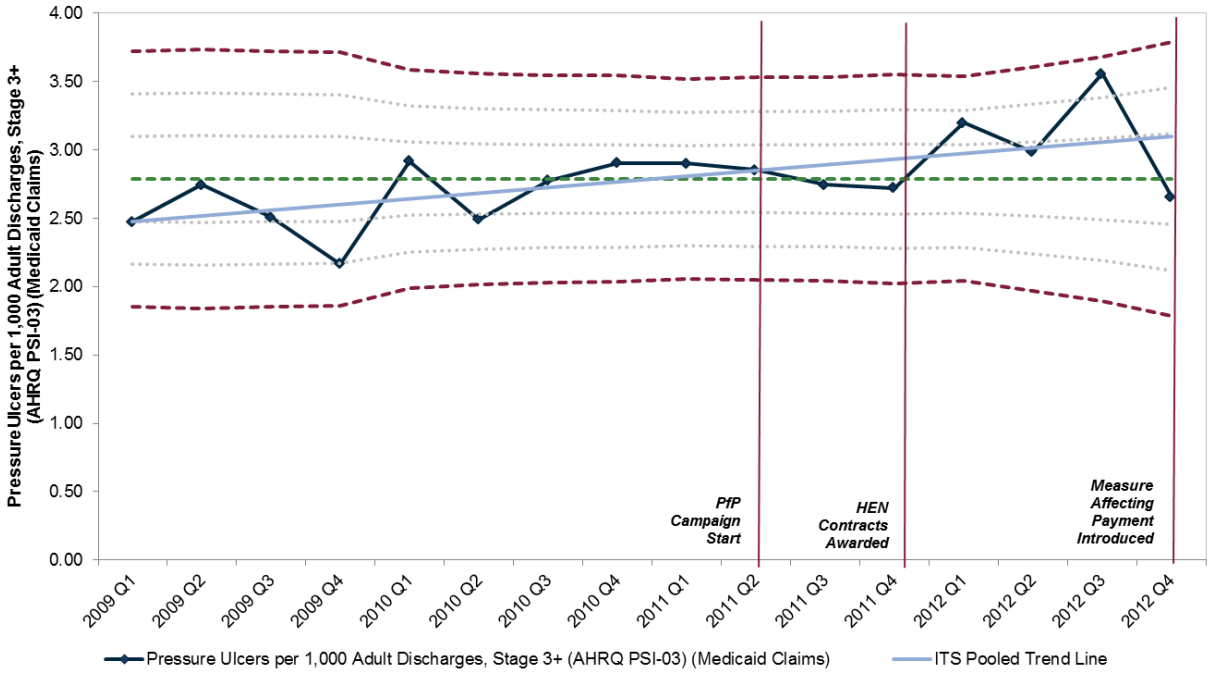
Source: Rates calculated by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS, based on Medicare FFS claims data. The Evaluation Contractor analyzed the claims data using a longer time series (from 2009) and adjusting for changing demographics over time. The analysis showed a similar pattern.

Note: Control limits and center line (U′ chart) were constructed using data from Q2 2011 to Q4 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two sigma limits. Calculations are based on Medicare FFS claims for all hospitals that reliably report POA status (≥ 95 percent of the hospital's diagnoses for a given quarter are accompanied by a valid code for POA) and that have the following characteristics: all hospitals paid under Medicare's IPPS, CAH, cancer hospitals, and Maryland hospitals. Data include between 721,477 and 846,909 discharges per quarter.

Pressure Ulcer per 1,000 Adult Discharges, Stages 3+ (AHRQ PSI-03) (Medicaid Claims) (Figure 2-44)

- No evidence of special cause variation.
- Evidence that the measure trended up without a significant break.
  - The average quarterly linear change was 0.045 per 1,000 adult discharges ($p < 0.001$).
- Rate decreased 4.13 percent between CY 2010 and Q4 2012.
- IPPS payment incentive related to pressure ulcers stages 3 and 4 became effective in October 2012.
- This measure may contain community-acquired as well as hospital-acquired pressure ulcers, since Medicaid claims do not contain information about whether the condition was present on admission.

Figure 2-44—Pressure Ulcer (Stages 3+) per 1,000 Adult Discharges (Medicaid Claims)



Source: Medicaid claims data for 17 states.
Note: Control limits and center line (U' chart) constructed using data from Q1 2009 to Q4 2012. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. Calculations are based on Medicaid claims data from 17 states (FFS data only in 14 states, MMC and FFS data in 3 states), assuming all diagnosis codes were not POA (Medicaid claims do not include POA data) and limiting the data to hospitals identified in the evaluation's roster of HEN-aligned and non-aligned hospitals. Regression-adjusted rates were obtained from a logit model that controlled for state dummies and patient-level controls (but not hospital fixed effects). Data include between 36,322 and 68,633 discharges per quarter. Medicaid data included in these analyses comprise claims for nondual eligible patients age 18 and older.

Pressure Ulcer per 1,000 Discharges (AHRQ National Scorecard) (Table 2-10)

- Possible evidence of change from the raw rates in this measure. This measure includes pressure ulcers of all-stages, including those of stage 1.
  - The 2012 rate is 2.18 percent lower than the 2010 rate. However, the 2013 rate is 19.35 percent lower than the rate in 2010.

Table 2-10—Pressure Ulcer per 1,000 Discharges (AHRQ)

2010 HAC Rates (per 1,000 discharges)	2011 HAC Rates (per 1,000 discharges)	2012 HAC Rates (per 1,000 discharges)	2013 HAC Rates (per 1,000 discharges)
40.31	40.41	39.43	32.51

Source: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html and Noel Eldridge, AHRQ Center for Quality Improvement and Safety, provided on August 28, 2015, for the final 2013 data.

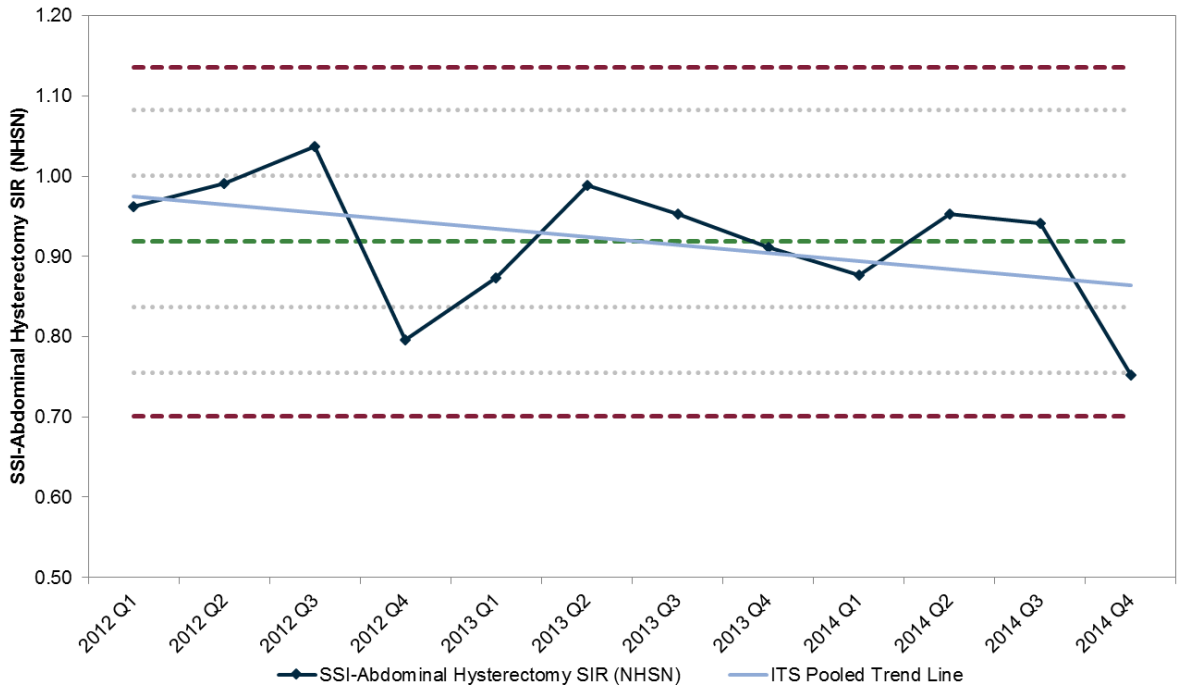
SSI

The AHRQ National Scorecard measure, which includes a range of surgeries, shows an 18.77 percent improvement between 2010 and 2013. The decline in the AHRQ measure between 2011 and 2013; however, is only 5.18 percent. By contrast, the SSI SIR from NHSN exhibits a 21.84 percent decline, although the trend fails to reach a magnitude for statistical significance and no structural breaks were detected in an ITS model.

SSI Abdominal Hysterectomy SIR (NHSN) (Figure 2-45)

- No evidence of special cause variation.
  - Control limits are temporary pending additional data.[2-13]
- No evidence of significant break or trend in the rate.
- Rate decreased 21.84 percent between Q1 2012 and Q4 2014.

Figure 2-45—SSI Abdominal Hysterectomy SIR (NHSN)



Source: NHSN. Data have 3,340 hospitals per quarter.
Note: Control limits and center line (X chart) constructed using data from Q1 2012 to Q4 2014. The dashed green line is the center line; the dashed red lines are the control limits; the closest dotted lines above and below the center line are the one-sigma lines; and the dotted lines just inside the control limits are the two-sigma limits. The data from NHSN were not provided with clustered standard errors that would account for the non-independence of discharges within the same facilities. Without the clustered standard errors, the SPC charts and ITS analysis performed are less conservative than would be the case if the correlation of outcomes within facilities were accounted for in the analysis.

[2-13] For SPC charts with fewer than 16 observations, the control limits are unstable, and thus final control limits cannot be calculated without additional data.

SSI per 1,000 Surgical Discharges (AHRQ National Scorecard) (Table 2-11)

- Evidence for improvement.
  - The 2013 rate is 18.77 percent lower than the 2010 rate.

Table 2-11—SSI per 1,000 Surgical Discharges*(AHRQ)

2010 HAC Rates (per 1,000 Surgical Discharges)	2011 HAC Rates (per 1,000 Surgical Discharges)	2012 HAC Rates (per 1,000 Surgical Discharges)	2013 HAC Rates (per 1,000 Surgical Discharges)
2.93	2.51	2.51	2.38

Source: The CDC estimates of SSIs based on NSHN data for 17 common surgical procedure types ("SCIP + 5") provided to AHRQ, forwarded by Noel Eldridge, November 13, 2014. Number of surgical procedures for those 17 procedure types was estimated by CDC from the Healthcare Cost and Utilization Project (HCUP) data for 2010-2012.

*Note that the percentage change in SSI rate is very similar regardless of whether AHRQ's per 1,000 discharges is used as the denominator or whether the denominator is specific to surgeries for the 17 procedure types (24.1 percent decrease using the AHRQ denominator versus 24.4 percent with the surgical discharges denominator shown above).

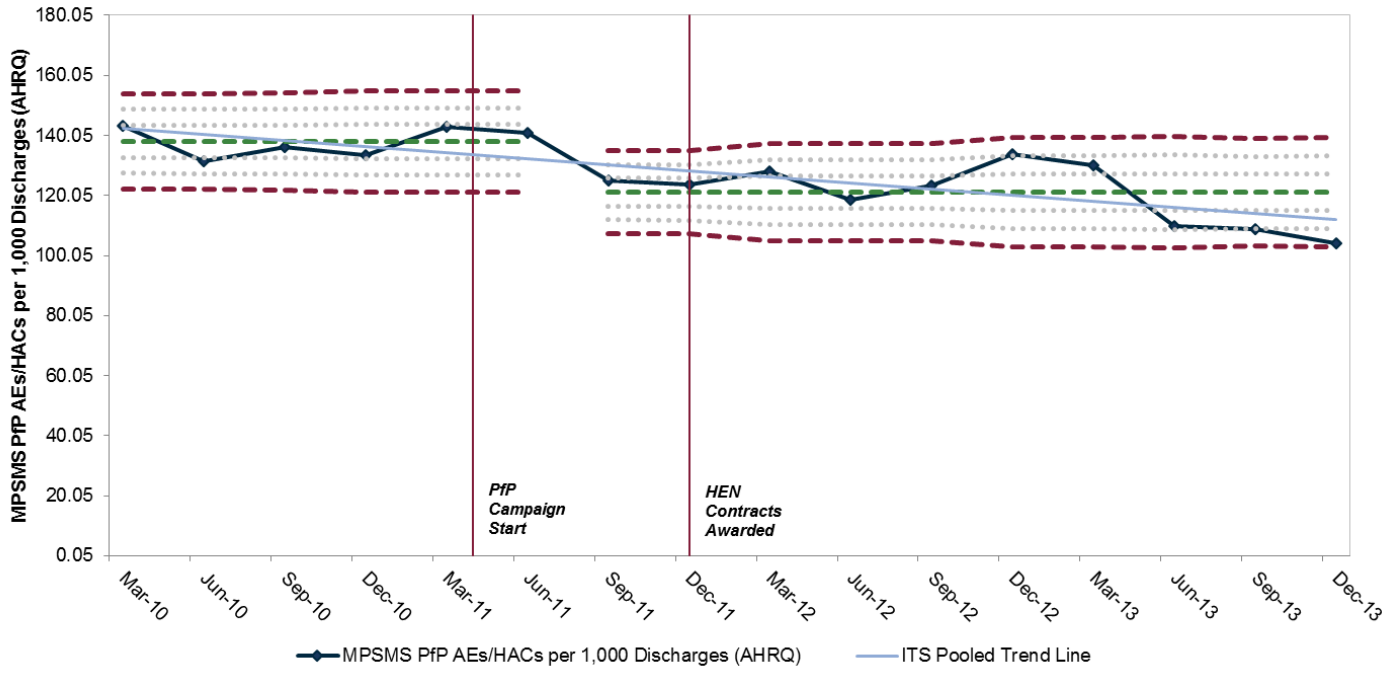
All Harm

The AHRQ National Scorecard measure, shows a 23.69 percent improvement between 2010 and 2013. The decline in the measure between 2011 and 2013; however, the trend fails to detect structural breaks were detected in an ITS model.

AHRQ National Scorecard HACs per 1,000 Discharges (AHRQ National Scorecard) (Figure 2-46)

- Evidence suggesting improvement in rate.
  - Shift in center line observed in Q3 2011.
- Evidence for a significant downward trend without a significant break.
  - The average quarterly linear change was -2.02 per 1,000 discharges (p>0.01).
- Rate decreased by 23.69 percent between Q1 2010 and Q4 2013.

Figure 2-46—AHRQ National Scorecard HACs per 1,000 Discharges (AHRQ)



Source: AHRQ National Scorecard.
Note: Center line and control limits (U' chart) for first phase were calculated with data between Q1 2010 and Q2 2011. Center line and control limits (U' chart) for second phase were calculated with data between Q3 2011 and Q4 2013. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

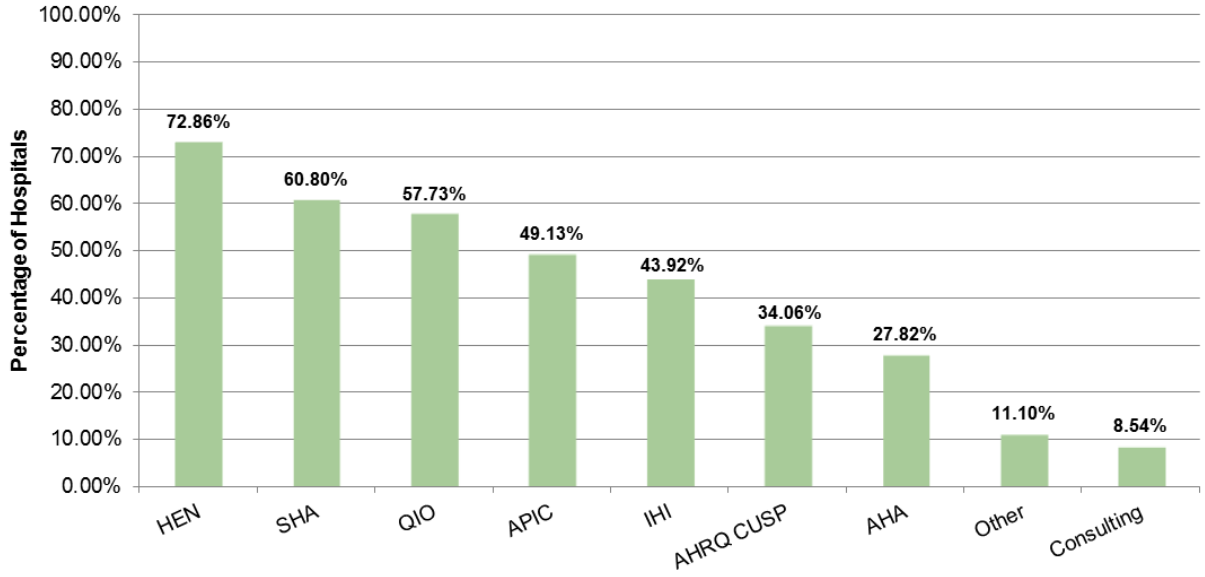
Concurrent PfP-Aligned and Parallel Efforts that May Have Contributed to the Reduction in Rates, in Addition to PfP

PfP was designed to actively align and encourage efforts underway by federal and non-federal organizations working toward patient safety goals (a.k.a. federal and non-federal partners), as described in Chapter 1. This section describes the extent to which hospitals that reported that their patient safety improvement and readmissions reduction actions were influenced by these other organizations and policies during the PfP timeframe. These activities occurred at the same time as the HEN activities and they cannot all be controlled for in these analyses. Due to this, the Evaluation Contractor cannot separately identify the effects of PfP from some of these other efforts.

Major Efforts at the National Level (Federal and Non-Federal Partner Efforts)

Hospitals responding to the Survey on Prevention of Adverse Events and Readmissions reported working with organizations other than HENs on patient safety improvement and readmissions reduction during 2012-2013 (Figure 2-47).[2-14] Over half of the hospitals indicated Quality Improvement Organizations (QIOs) as organizations they worked with on patient safety and readmissions improvement, and nearly half (49.13 percent) worked with the Association of Professionals in Infection Control and Epidemiology (APIC). In addition, the Institute for Healthcare Improvement (IHI) and AHRQ Comprehensive Unit-Based Safety Program (CUSP) programs worked with 43.92 percent and 34.06 percent of hospitals, respectively.

Figure 2-47—Percentage of Hospitals that Worked with Various Types of Organizations on Patient Safety During 2012-2013



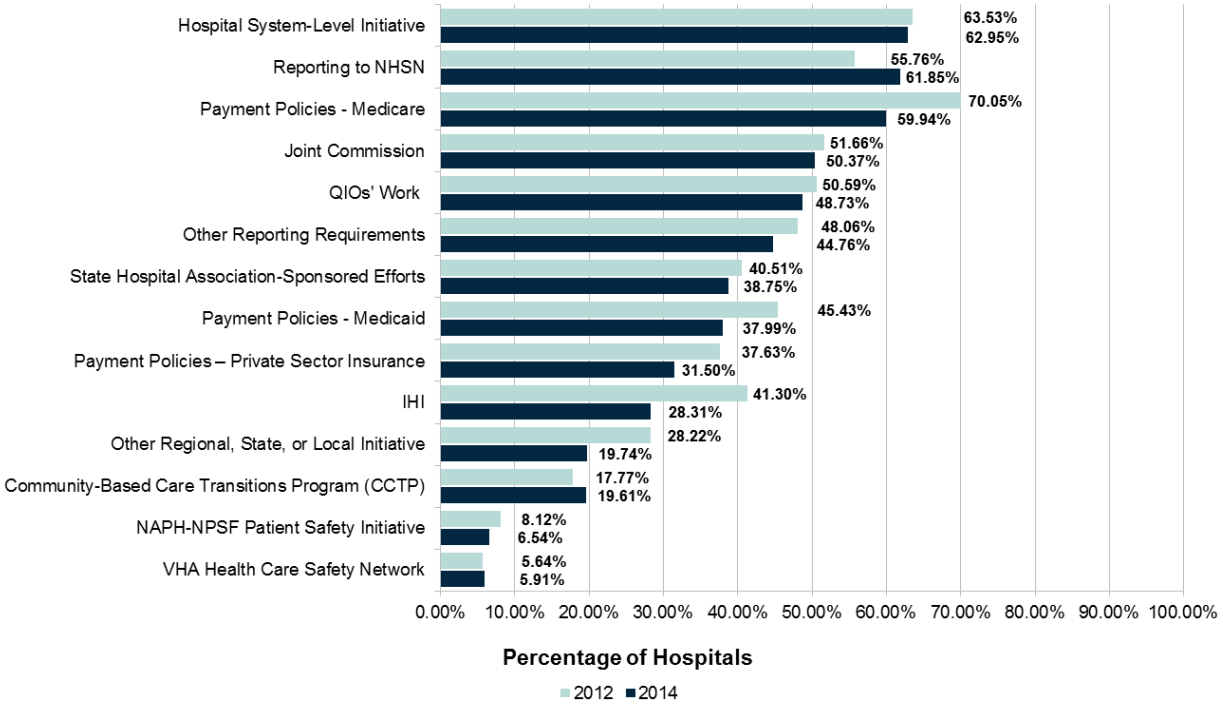
Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2014.

2-14 The survey was a national sample survey of all U.S. short-term acute care hospitals (HEN-aligned and non-HEN-aligned) by the Evaluation Contractor. This finding is from the second round of the survey in 2014 (see Appendix C for more details on the method). In the survey, state hospital associations (SHAs) and the American Hospital Association (AHA) are related to PfP but were listed separately from HENs to facilitate recognition by hospitals that may not know that their SHAs have been working on patient safety through the AHA/HRET HEN.

Hospitals reported that a variety of policies and programs beyond the HENs had influenced them to take actions to prevent harm and readmissions during the past 12 months, in both the 2012 and 2014 rounds of the survey (Figure 2-48). Between 51 and 64 percent of all the nation's hospitals reported the following as influencing their harm reduction actions: hospital system organizations, the CDC's NHSN reporting system (and presumably associated Medicare requirements to report to it), Medicare payment policies, the Joint Commission's accreditation requirements, and QIOs' work.

The influence of Medicare payment policies, IHI, and Medicaid payment policies on hospital actions was less in the year prior to the 2014 survey than in the year prior to the 2012 survey, though still a majority of hospitals were influenced by Medicare payment policies, while reporting to NHSN was more influential in 2014.

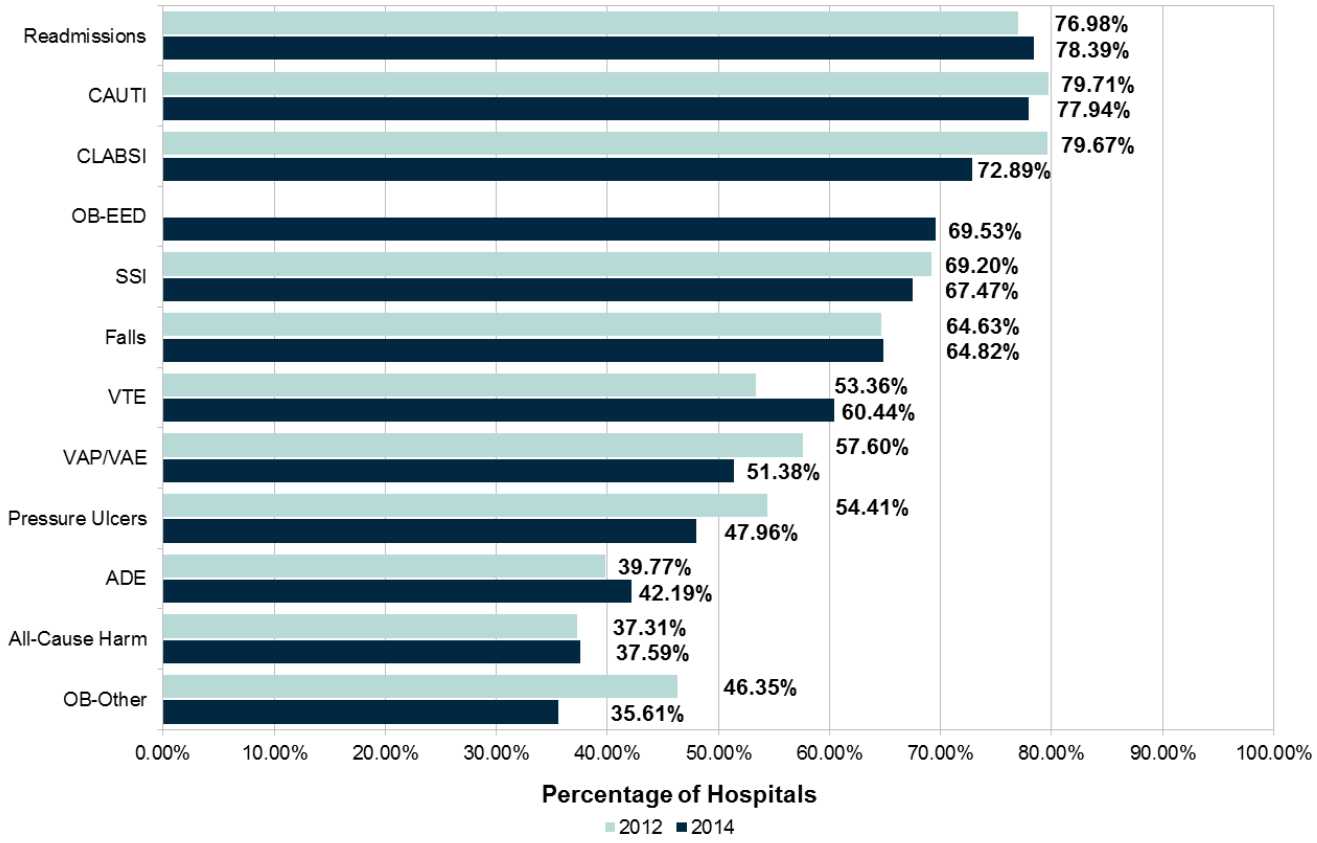
Figure 2-48—Percentage of Hospitals Citing Other Influences Beyond HENs on Actions to Prevent Harm and Readmissions



Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and 2014.
Note: The bars shown for State Hospital Association-Sponsored Efforts may reflect PfP-related activity, since many state hospital associations either subcontracted with the AHA/HRET HEN, or served as HENs themselves. Not all of the influence shown here is likely to reflect HEN influence, since hospital associations sometimes undertook patient safety-related projects funded by other sources, a small number of states were neither HENs nor subcontracted to AHA/HRET HEN, and many hospitals belonged to a HEN that was not their state hospital association, such that their state hospital association influence was not the same as the influence from their HEN.

These other efforts did not influence all adverse event areas equally. Over two-thirds of all hospitals reported these factors had influenced their work on readmissions, CAUTI, CLABSI, OB-EED, and SSI in the year prior to the 2014 survey, and between half and two-thirds reported these factors had influenced their work on falls, VTE, and VAP/VAE (Figure 2-49).

Figure 2-49—Percent of Hospitals Reporting on Each Area that were Affected by Policies, Programs, and Entities Other than HENs



Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and 2014.
Note: OB-EED was not included in the 2012 survey. Given the specific survey instrument used, it was not possible to separate the influence of state hospital associations (some of which was HEN-related) from the influence of the other factors listed in Figure 2-48. Therefore, some of the influence shown here is likely to reflect HEN influence through state hospital associations.

Major national-level federal and non-federal policy and program efforts are listed by PfP focus area in Appendix E.

Insights from Site Visits with 12 Hospitals

Visited hospitals indicated that many factors influenced their quality improvement processes.[2-15] These included both external factors (such as the policy initiatives described above) and internal hospital factors. Hospitals that were included in our site visits indicated that there were many policies, as well as other initiatives, that affected their quality improvement efforts. These included many of the policies that were included in our survey, such as CMS and state Medicaid payment policies, QIO projects, public reporting initiatives, as well as others such as health information technology (HIT) meaningful use policies.

In addition, there were many external local factors that played an important part in forming hospitals' quality improvement initiatives that were not associated with government policy. These included participation in a network of peer hospitals, projects with patient safety organizations, private foundations, and physician specialty society initiatives.

The site visits also revealed the importance of "internal factors" in the hospitals' quality improvement efforts. The most common internal factors influencing visited hospitals' change concerned strategy (6 hospitals), leadership (4 hospitals), and structure (3 hospitals). As an example related to strategy, 3 hospitals had committed to maintaining Magnet status for excellence in nursing in an effort to attract patients. This commitment involves participating in NDNQI and tracking and seeking to improve on National Quality Forum (NQF)-endorsed falls and pressure ulcers measures. An example related to leadership is hiring a new chief medical officer (CMO) and quality director who brought the experience and determination to change the organization's culture. An example related to structure is having an integrated physician group, so that the organization can implement fully aligned harm prevention initiatives with buy-in from all the relevant physicians.

QIO/PfP Trends

A critical question in the evaluation of the HEN impact during the PfP campaign is to what extent were observed reductions in patient harms associated with other quality improvement initiatives that were occurring at the same time as PfP. If the impacts of the PfP campaign were enhanced by exposure to additional intervention programs, then hospitals exposed to multiple programs might be expected to exhibit greater reductions in patient harms than those exposed to individual programs alone. One program, operating on a national scale, whose impact may overlap with the PfP campaign and which focused on several of the same harm areas (e.g., hospital-acquired infections and readmissions) was CMS' QIO program.

To examine the question of patient harms reductions in PfP-aligned hospitals versus those working with a QIO, the Evaluation Contractor performed three distinct analyses. First, the outcome trends for hospitals working only with PfP were compared to trends for hospitals working only with a QIO. This analysis shed light on whether the trends for the two groups can be distinguished from one another, as the ability to distinguish between the two groups is critical for disentangling the potential impacts attributable to each campaign. Second, the outcome trends for hospitals participating with both PfP and a QIO were compared to those that worked with either campaign individually or with neither campaign. This analysis provided an understanding of whether hospitals working with both campaigns exhibited greater reductions in patient harms than those that only worked with one or neither campaign. Finally, the analysis compared the trends for hospitals that worked with either one, or both PfP and a QIO against the trends for hospitals that worked with neither campaign. This

[2-15] A stratified random sample of hospitals was selected for the visits, to ensure a mix in terms of level of engagement with PfP (from non-alignment to full engagement), hospital type (CAH/non-CAH) and size, urban/rural location, and geographic location. See Appendix C for more detail on the method.

analysis provided insight into the relative relationship between campaign affiliation and outcome trends across the different participation groups.

The Evaluation Contractor compared the trends in hospital-level patient safety outcomes for PfP-aligned hospitals with those aligned with CMS-funded QIO programs in four states: Arizona (AZ), California (CA), Florida (FL), and Ohio (OH).[2-16] The analyses assessed whether the PfP campaign affiliation was associated with uniquely identifiable trends in readmission rates and three PSIs, as shown in Table 2-12.[2-17] The period of analysis was from Q1 2010 through Q1 2014 for readmission rates and from Q2 2011 through Q1 2014 for the three PSIs. All outcome measures were calculated using Medicare claims data.

Table 2-12—Measure Rates Examined

AEA	Measure
CRBSI	CRBSI rates (PSI-07)
Pressure Ulcers[2-18]	Pressure ulcer rates (PSI-03)
VTE[2-19]	VTE Postoperative PE or DVT per 1,000 surgical discharges (PSI-12)
Readmissions	30-day, all-cause readmissions

Source: Medicare claims data.

Methodology

The population of hospitals that submitted Medicare claims data between Q1 2010 and Q1 2014 in the four states numbered 884. The hospitals logically fell into one of the four "treatment groups": those that (1) participated in the PfP campaign only, (2) participated with the QIO-only, (3) participated with both the PfP and QIO, or (4) participated with neither the PfP nor the QIO program. The number of hospitals in each treatment group is provided in Table 2-13. Descriptive statistics for the outcomes rates and hospital characteristics are included in Appendix E.

Table 2-13—Hospital Treatment Groups

Treatment Group	Number of Hospitals
PfP-Only	312
QIO-Only	75
PfP and QIO	184
Neither PfP nor QIO	313
Total	884

Source: The Evaluation Contractor's tabulations of QIO participation list form the QIO 10th Statement of Work (SoW) and PfP November 2014 hospital lists.

2-16 The Evaluation Contractor was only able to obtain data for AZ, CA, FL, and OH for this analysis.

2-17 Rates for PSI-03 and PSI-07 are per 1,000 medical and surgical discharges. The PSI-12 rate is per 1,000 surgical discharges. The 30-day all-cause readmissions rate includes readmissions to the same facility within 30 days of an index stay. All rates were calculated by the Evaluation Contractor using Medicare claims.

2-18 AHRQ's PSIs.

2-19 Ibid.

To determine if differences between groups were statistically significant, the Evaluation Contractor estimated hierarchical generalized linear models (HGLMs) of the trends within each analysis group. Specifically, these HGLMs allow each hospital to have its own average level and trend over time. The hospital-specific levels or trends can then be assessed to determine if hospitals in one group exhibited significantly different levels or trends in outcomes from hospitals in another group. If the results do not identify statistical differences in the trends of the different campaign participant groups, then the evidence will suggest that the relative impacts of the PfP and QIO campaigns cannot be disentangled from one another.

Results

Identifying Differences between PfP-Only and QIO-Only Hospitals

Table 2-14 provides the average difference between quarterly change in the rates for PfP-only and QIO-only hospitals over the study period. For example, the CRBSI infection trend for PfP-only hospitals declined by 2.30 percent less than QIO-only hospitals; however, this difference was not statistically significant. None of the measures evaluated showed a statistically significant difference in the trends between the PfP-only group and QIO-only group. Although the trends for PfP-only were generally higher than QIO-only, these results were indistinguishable from statistical noise. This means that the HGLMs were not able to differentiate between the improvements made by the QIO-only hospitals and those made by the PfP-only hospitals.

Table 2-14—Hierarchical Generalized Linear Model Solution for Fixed Effects Incremental Impact of PfP-Only Hospitals Compared to QIO-Only Hospitals

Treatment Group	CVC Infection Count (per 1,000 discharges)	PE/DVT Count (per 1,000 discharges)	Pressure Ulcer Count (per 1,000 discharges)	30-Day All-Cause Readmissions Count
PfP-Only	2.30%	-1.63%	5.08%	0.10%

Source: HSAG's analysis of QIO participation and Medicare claims data.
Note: Estimates are solutions for fixed effects. The full regression model results tables can be found in Appendix E.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Identifying Differences in Trends for Hospitals Aligned with Both or Neither Program

In addition to evaluating the differences in trends between PfP-only hospitals and QIO-only hospitals, the Evaluation Contractor compared the trends in hospitals that participated in both PfP and QIO against each of the other three treatment groups. Each number in Table 2-15 represents the difference in the trends for the listed treatment group compared to hospitals aligned with both the PfP and QIO. For example, the quarterly trend in PE/DVT for the PfP-only aligned hospitals was 1.36 percent smaller than hospitals in both the PfP and QIO. This suggests that the PfP-only hospitals saw a significantly greater rate of improvement (at the 0.10 level) for PE/DVT than hospitals in both the PfP and QIO. However, hospitals that participated in either the QIO or PfP alone, or that did not participate in either program at all, generally achieved a smaller average quarterly rate of improvement (indicated by positive results in Table 2-15) than hospitals that participated in both the PfP and QIO programs.[2-20]

[2-20] An important caveat to this analysis is that it is not possible for the Evaluation Contractor to know what these rates would have been in the absence of either intervention program. It is possible that without the PfP or QIO programs, those hospitals could have seen an increase in average rates of patient harm rather than a decrease.

The results were mixed when comparing hospitals that worked only in PfP to those that participated in both PfP and the QIO. Those participating in PfP-only achieved a significantly smaller rate of improvement for two measures—CVC infections and pressure ulcers—but a significantly greater rate of improvement for the PE/DVT measure. The difference in the rates of improvement for the 30-day, all-cause readmissions measure between the PfP-only group and PfP and QIO group was not significant.

Hospitals that worked only with the QIO exhibited a smaller average quarterly rate of improvement than those that participated in both the PfP and the QIO for all measures except for the 30-day, all-cause readmissions measure. However, none of these differences were statistically significant.

Table 2-15—Hierarchical Generalized Linear Model Incremental Impact of Each Treatment Group Compared to PfP and QIO Hospitals

Treatment Group	CVC Infection Count (per 1,000 discharges)	PE/DVT Count (per 1,000 discharges)	Pressure Ulcer Count (per 1,000 discharges)	30-Day All-Cause Readmissions Count
Neither PfP nor QIO	3.98%	0.09%	6.51%*	0.07%
PfP-Only	3.98%**	-1.36%*	7.66%***	-0.01%
QIO-Only	1.97%	0.34%	1.92%	-0.11%

Source: HSAG's analysis of QIO participation and Medicare claims data.
Note: Estimates represent the average quarterly change in rates. The full regression model results tables can be found in Appendix E.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, the Evaluation Contractor further compared the trends in hospitals that participated in neither PfP nor QIO against each of the other three treatment groups. Each figure in Table 2-16 represents the difference in trends for the listed treatment group compared to hospitals aligned with neither the PfP nor the QIO. For example, the quarterly trend in PE/DVT for PfP-only hospitals was 1.48 percent lower than that of hospitals aligned with neither program, which suggests a greater improvement in the rate. However, none of the differences in measures evaluated for the PfP-only and QIO-only hospital groupings was statistically significant at the 0.10 level. As illustrated in both Table 2-15 and Table 2-16, hospitals in both PfP and QIO had a significantly greater improvement in the trend for pressure ulcers compared to hospitals in neither PfP nor QIO.

Table 2-16—Hierarchical Generalized Linear Model Incremental Impact of Each Treatment Group Compared to Neither PfP nor QIO

Treatment Group	CVC Infection Count (per 1,000 discharges)	PE/DVT Count (per 1,000 discharges)	Pressure Ulcer Count (per 1,000 discharges)	30-Day All-Cause Readmissions Count
PfP and QIO	-3.98%	-0.09%	-6.51%*	-0.07%
PfP-Only	-0.18%	-1.48%	0.83%	-0.07%
QIO-Only	-1.88%	0.26%	-3.69%	-0.17%

Source: HSAG's analysis of QIO participation and Medicare claims data.
Note: Estimates represent the average quarterly change in rates. The full regression model results tables can be found in Appendix E.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The purpose of this analysis was to provide insight into whether the results of the PfP campaign could be distinguished from those of the other QIO programs going on in tandem. The aim was to assess the hospital-level trends in selected patient outcomes and determine whether it was possible to distinguish between the trends of hospitals that participated in the PfP campaign and those that worked with a QIO. Importantly, the trends in measure rates did not differ significantly between the group of hospitals that participated in neither PfP nor a QIO alone. Thus, the data did not provide enough signal to effectively distinguish between the two groups based on the intervention in which a hospital participated. Any observed differences in the trends are more likely to be a function of random fluctuations in the data than of any meaningful difference in program impact. By contrast, hospitals that worked with both PfP and a QIO were significantly different from those hospitals that worked only with the PfP. There was some evidence that among hospitals that worked with the PfP campaign, additional participation in the state QIO program was associated with larger reductions in patient harms.

Limitations

As with any analysis, there are limitations to the inferences that can be drawn. Perhaps most importantly, the results presented in this analysis do not speak to the impact of either the PfP campaign or the work performed by the QIO on rates of patient harms. The focus of the analysis was placed on whether it was possible to differentiate between the trends exhibited by hospitals participating with each quality improvement initiative.

Another limitation of the analysis was that the data are only generalizable to hospitals in the states of CA, AZ, FL, and OH. The data available to the Evaluation Contractor were only from hospitals in these states and cannot provide any insights on the differences between PfP-only and QIO-only hospitals in other states.

A third limitation to note is that the outcome measure rates used for this analysis are not risk-adjusted for differences in patient demographic and clinical diagnosis characteristics. Thus, if patients who were more susceptible to experiencing specific patient harms are differentially allocated to either PfP or QIO hospitals, the lack of significant group differences may mask differences in risk-adjusted rates of harm. Furthermore, the extent to which QIO hospitals were working on each of the harm areas included in this analysis is unclear. Therefore, the results may be biased to the extent that there was variation in the hospital participation in reducing these patient harms between the PfP and QIO campaigns.

A final limitation was that the analysis is not able to account for the self-selection bias inherent in studying the PfP and QIO hospitals. For both campaigns, hospitals participated because leadership chose to become affiliated with the campaign. Such self-selection processes often create fundamental differences between those that participated and those that did not. It is therefore difficult to determine if the findings are associated with differences in campaign affiliation, or differences in other characteristics related to the processes of selection into a specific quality improvement campaign.

3. PfP Learning Community's Work Toward Reduction in Harms

As previously discussed, the Partnership for Patients (PfP) campaign centered on the development of a shared learning community to assist in the spread of best practices in the PfP focus areas. Hospital Engagement Networks (HENs) were a key component of this learning community and represented the largest part of the Center for Medicare & Medicaid Innovation (CMMI) investment. This learning community had to effectively spread best practices to hospitals on a very large scale and this had to result in a widespread response by the HEN-aligned hospitals. In turn, to result in improved outcomes, the changes made at the hospital level had to influence the outcomes as expected.

This chapter discusses the PfP learning communities' work toward a reduction in patient harms, and it includes the following sections:

- The spread of best practices by the HENs.
- Changes made, based on the spread of best practices, by engaged hospitals.

The data used in this chapter came from the following sources (see Appendix C for details):

- Interviews in fall 2014 with all HENs and with state hospital associations (SHAs) that were part of the AHA/HRET HEN (referenced below as HEN interviews, fall 2014).[3-1]
- Interviews in summer 2014 with all HENs (referenced below as HEN interviews, summer 2014).
- Interviews in winter/spring 2015 with representatives from 19 non-federal partners, that is, organizations outside the federal government working in patient safety during PfP (referenced below as non-federal partner interviews, winter/spring 2015).
- Reviews of federal partner organizations' written summaries of their contributions to PfP, provided November 2014 (referenced below as federal partner summaries, November 2014).
- Site visits in spring 2015 to 12 hospitals, both HEN-aligned and non-HEN aligned hospitals (referenced below as site visits, spring 2015).
- Two surveys of hospitals, which were designed to show changes over time, for which background is provided in Table 3-1 and Table 3-2 (referenced below by the names of the two surveys). Note that the second survey (Table 3-2) is not generalizable beyond the respondent group, but rather was designed to document the extent of participation with HENs and response to HEN activities within as many from the universe of HEN-aligned hospitals as could be persuaded to respond.
- HENs' Z-5 tables, provided monthly by HENs to the Centers for Medicare & Medicaid Services (CMS) and the Evaluation Contractor; contents include HENs' report on each aligned hospital's status of implementing five specific patient family engagement (PFE) processes and four leadership processes (referenced below as HEN Z-5 tables).[3-2]

3-1 As described in Chapter 1, one of the 26 HENs (AHA/HRET) served 37 percent of the hospitals aligned with PfP HENs. AHA/HRET operated through 31 SHAs, which both leveraged AHA/HRET's national-level infrastructure, and used PfP subcontract funding to directly assist hospitals in their state.

3-2 Hospital participating (Z-5) spreadsheets were submitted by HENs and included a Z-5 scale with hospital-specific status indicators, as well as questions on PFE and leadership activities. These data were submitted using a template that can be found on the Community of Practice website: http://www.healthcarecommunities.org/DesktopModules/Bring2Mind/DMX/Download.aspx?PortalId=3&TabId=1609&EntryId=43829.

Table 3-1 presents background information about the Survey on Prevention of Adverse Events and Reduction of Readmissions conducted in spring 2012, and repeated in spring 2014.

Table 3-1—Background on Survey on Prevention of Adverse Events and Reduction of Readmissions	
Fielding Periods	Spring 2012, repeated spring 2014
Mode	Web-based
Sample Type	Stratified random sample of short-term acute care hospitals, both HEN-aligned and non-HEN-aligned, excluding Veterans Health Administration (VA) and military hospitals.
Sample Size and Response Rate **Spring 2012** **Spring 2014** **Used in this report**	 2,452, of which 1,719 completed the survey (70 percent). 1,706, of which 1,136 responded (67 percent).[3-3] 1,136 hospitals responding to both surveys (47 percent).
Topics	• Participation and experience with PfP. • Improvement efforts and observed changes in outcomes. • Influences on improvement efforts and outcomes. • Hospital culture and patient care practices related to patient safety and readmissions reduction.
Sample Weighting for Analysis	All 88 children's hospitals were purposely selected into the sample, so these were re-weighted to be proportionally represented in the analyses.
Non-Response Weighting	Non-response weights were used to adjust for the different characteristics of responding vs. non-responding hospitals, separately for HEN-aligned and non-HEN-aligned hospitals, and overall. A bias analysis confirmed that the likely bias from differential non-response among hospital types was nearly eliminated by the weights (Appendix C). There may remain bias from unmeasured factors, such as motivation to pursue patient safety activities.
Precision	A yes/no question to all respondents with an even split of responses can be interpreted as true + or – 3 percentage points. All survey results expressed as percentages are weighted estimates. Each has its own confidence interval, but for readability, these are not presented in the report, although they are accounted for in the statistical test results that appear throughout.

[3-3] The spring 2014 sample excluded several hospitals that had merged or closed since responding to the 2012 survey.

Table 3-2 presents background information about the Survey on Participation in Patient Safety Activities conducted in January 2015 through March 2015.

Table 3-2—Background on Survey on Participation in Patient Safety Activities	
Fielding Period	January through March 2015
Mode	Web-based
Sample Type	All HEN-aligned hospitals as listed by HENs
Sample Size and Response Rate	3,389, of whom 2,432 completed it (72 percent)
Topics	• Participation in specific types of HEN activities. • Operational changes made due to participation in HEN activities, by adverse event area (AEA). • Operational changes made, but not due to participation in HEN activities. • Space for comments on hospital's experience with PfP.
Sample and Non-Response Weighting for Analysis	No weighting
Generalizability	The results cannot be generalized beyond the respondents; rather than sampling, the entire population of interest (HEN-aligned hospitals) was the target for the survey; it would be expected that hospitals that did not respond were less involved in PfP and would have different views.

Spreading of Best Practices Through HENs

This section summarizes what the HENs did and the factors that affected their ability to spread best practices. Unless otherwise noted, text is based on qualitative analysis of interviews with all 26 HENs in fall 2014. Quotations from HENs and hospitals that participated in a set of 12 site visits are used to illustrate points derived from the qualitative analysis, and other data sources are incorporated where appropriate and specifically noted.[3-4]

Services Provided by HENs to Hospitals

HENs' implementation of PfP involved an extensive set of activities. As part of the PfP learning community, HENs implemented a wide-ranging set of activities to support hospital-level harm reduction (Table 3-3). Most HENs engaged in all of these strategies, using multiple strategies to address the diverse needs of each PfP focus area. HENs' implementation strategies were largely consistent with several key principles of effective implementation identified through eight reviews of the literature regarding quality improvement initiatives and the adoption of evidence-based practices in health services settings (analysis provided in Appendix E). These strategies employed by the various HENs are summarized in Table 3-3.

[3-4] Quotations were selected to illustrate themes that emerged from the qualitative analysis. However, no quotation may be considered representative of any but the quoted organization.

Table 3-3—HEN Strategies to Spread Best Practices for Patient Safety—The What and How	
Overarching Strategy	**Tactics**
Engage hospitals and key staff	• Engage hospital leadership • Engage frontline staff and clinicians • Consult with hospitals on uses in PFE harm reduction • Create new partnerships or build on existing partners to carry out harm reduction activities • Engage community stakeholders and encourage hospital partnership at the local level • Align with concurrent or existing initiatives (e.g., public reporting requirements, Medicaid payment policy changes)
Change processes of care	• Develop, select, and disseminate interventions • Create opportunities for shared learning and networking • Monitor performance and provide feedback • Consult and coach on processes of care • Adapt interventions to local needs and circumstances • Target strategies to specific hospitals based on performance • Test and spread interventions • Standardize interventions used within hospitals or across the HEN network
Build capacity for safety-across-the-board	• Educate hospitals and their staff • Provide skills training to hospitals and their staff (for example, Lean/Six Sigma) • Support development of data infrastructure and measurement

Source: The Evaluation Contractor's interviews with all 26 HENs, fall 2014.
Note: See Appendix C for details on the methodology.

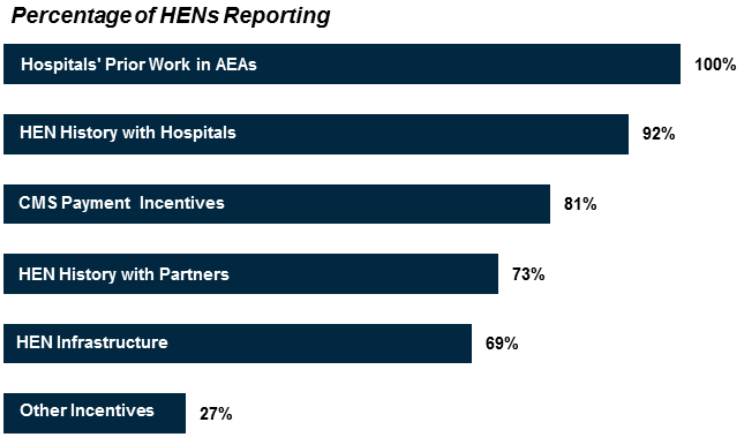
Factors Affecting HENs' Ability to Spread Best Practices

A variety of factors influenced HENs' ability to spread best practices. First, research has identified a variety of important determinants of program implementation effectiveness and successful scale up and spread of best practices in healthcare.[3-5] Applying these research-based frameworks to the analysis of the PfP campaign, several organizational (HEN- and hospital-level) and external factors emerged as having an effect on HEN efforts to spread best practices for harm reduction in their aligned hospitals. Facilitators and barriers to successful spread are identified in Figure 3-1 and Figure 3-2, respectively. Second, HEN feedback regarding key features of PfP also identified aspects of the campaign that were facilitators of progress, including PfP's bold aims, support contactors' work, interim targets and assessments, focused "pushes" within the campaign, and emphasis on partnerships. In addition to HENs, a total of 31 state hospital associations (SHAs) participated in PfP through the AHA/HRET HEN and reported similar influences on their harm reduction work. Third, factors related to work in specific focus areas also affected HEN efforts to reduce harm in some areas.

[3-5] More detail about research-based frameworks for program implementation and scale up and spread of best practices and how these best practices were applied is found in Appendix C.

Figure 3-1 provides the percentage of HENs' agreement as to facilitators to spread of best practices.

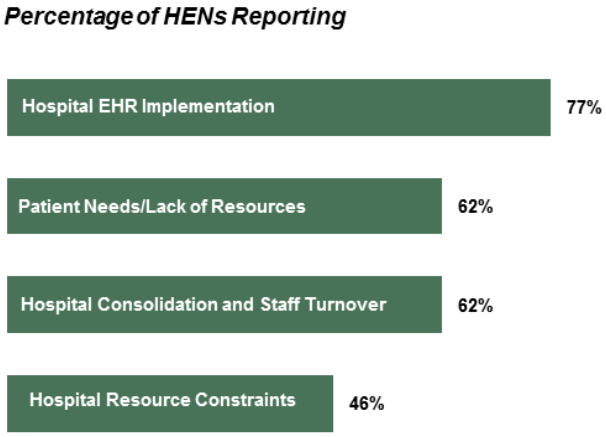
Figure 3-1—Facilitators of HEN Spread of Best Practices



Source: Evaluation Contractor interviews with 26 HENs, fall 2014.

Figure 3-2 presents the percentage of HENs' agreement as to barriers to spread of best practices.

Figure 3-2—Barriers to HEN Spread of Best Practices



Source: Evaluation Contractor interviews with 26 HENs, fall 2014

Facilitators

Organizational and External Factors

HENs capitalized on prior work focused on PfP target areas and existing infrastructure to support their harm reduction efforts during the PfP campaign. Many (18) of the HENs took advantage of existing organizational resources including strong clinical leaders, topic area experts, and advisory councils, expanding their work through participation in PfP. All HENs reported prior work in their hospitals to reduce target areas of harm. Half of the HENs (13 HENs) were working with their aligned hospitals in five or more PfP target areas prior to the campaign, although typically this work involved a smaller subset of hospitals than they engaged during PfP. HENs' prior work with hospitals focused on some harm areas more than others. Ten or more HENs reported working with their hospitals to reduce catheter-associated urinary tract infections (CAUTI), central line-associated blood stream infections (CLABSI), surgical site infections (SSI), and readmissions prior to PfP. Six or seven HENs worked with their hospitals prior to PfP on ventilator-associated pneumonia (VAP), pressure ulcers, or falls, while adverse drug events (ADE), other obstetrical adverse events (OB-Other), and venous thromboembolism (VTE) received less focus, with four or fewer HENs reporting earlier work in these areas.

Trust and relationships with hospitals and partners developed through prior work allowed HENs to achieve strong buy-in and commitment at the start of PfP. Benefits of HENs' history working with aligned hospitals (24 HENs) and partners (19 HENs) prior to PfP included established working relationships, access to data, and lessons learned from previous patient safety improvement efforts that provided a strong foundation for collaboration and development of successful strategies to tackle new areas.

The CMS payment incentives generated a sense of urgency and accelerated hospitals' adoption of safety improvements. Many HENs (21 HENs) reported that the CMS Hospital Readmissions Reduction Program (HRRP), the Hospital-Acquired Condition (HAC) Reduction Program, and the Hospital Value-Based Purchasing (VBP) Program had a positive impact on driving change by motivating hospitals and helping focus patient safety improvement efforts. However, some HENs (8 HENs) felt the overall impact of payment penalties was limited in some ways. First, non-inpatient prospective payment system (IPPS) hospitals and non-Medicare admissions are not eligible for payment incentives. Trend data shows harms were reduced within Critical Access Hospitals, not subject to CMS payment incentives (see Appendix E), so other factors in addition to payment incentives were clearly at work. Second, hospitals' responses to payment incentives varied. Two HENs described the challenge hospital leaders faced to justify investing resources in efforts to reduce readmissions from a financial standpoint, given a greater loss in revenues from reducing readmissions than from the penalty under the HRRP.

> *"Payment was a factor. When they talked about reducing payment for readmissions, that was a huge incentive. Lots of hospitals can't afford to lose those payments."*–a visited hospital

> *"Let's not underestimate the impact that payment policy reforms made on reductions. They really put the spotlight on harm and strategies to reduce harm and readmissions."*–a HEN

Alignment of PfP goals with state Medicaid policies, state mandatory reporting, and commercial payer incentive programs provided incentives to improve care in hospitals and patient populations that are not eligible for CMS payment incentive programs. A number of HENs (7 HENs) reported that alignment with other state and commercial programs helped increase awareness and commitment to readmissions and harm events targeted by PfP. Medicaid nonpayment for obstetrical early elective delivery (OB-EED) helped engage hospitals and encourage implementation of hard-stop policies. Quality goals of commercial payer pay-for-performance programs aligned with PfP aims provided levers for HENs to engage hospitals, including rural and critical access hospitals (CAHs), and increase commitment to improve patient safety.

State mandatory reporting programs initiated prior to PfP, particularly related to hospital-acquired infections (HAIs) (such as CAUTI, CLABSI, and SSI), as well as pressure ulcers, falls, and VAP, also helped engage key leaders.

PfP Campaign-Related Factors

Positive feedback on key PfP design features suggests many of these, too, were facilitators of progress (see Appendix E for details):

- **National-level support contractors provided helpful content**. HENs reported that work by all three support contractors facilitated their progress (21 HENs), citing high-quality content and speakers for the learning events, high-quality PFE master classes, and feedback reports that allowed them to compare their progress to that of other HENs.
- **Bold aims were reported to generate urgency and help engage hospitals**. By setting bold aims of a 40 percent reduction in preventable inpatient harm and a 20 percent reduction in readmissions, PfP hoped to generate more energy and excitement among all stakeholders than would be the case with more modest ambitions. Discussions with HENs suggested the boldness of these aims was a source of motivation, creating urgency and focus around efforts to improve patient safety. HENs reported that the aims were also critical to generating engagement from hospitals and hospital leadership.

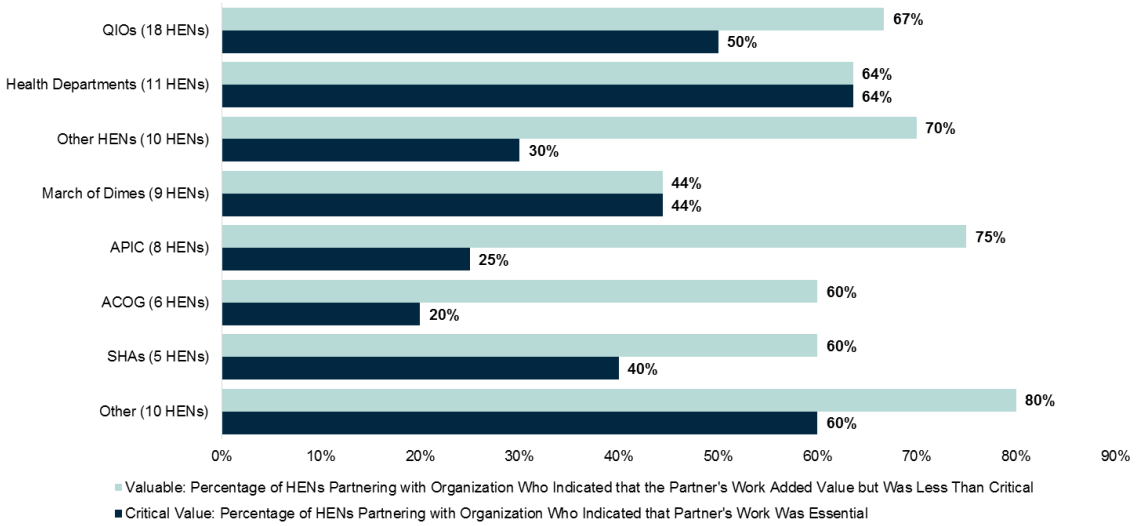
> *"We were already doing tracking before PfP. Before, though, we didn't have a goal. PfP gave us that."*–a visited hospital

- **Interim targets and related assessments allowed HENs to gauge their progress toward the goals**. In addition, HENs reported that the interim goals assisted them with hospital engagement because hospitals viewed them as more feasible short-run goals than the 40/20 goals.
- **Focused "pushes" within the campaign added urgency to work on OB-EED, readmissions, and CAUTI**. As described in Chapter 1, these pushes entailed asking HENs and hospitals to focus intensely on reducing a specific adverse event and report progress back to CMS. Most HENs (18 of 25 that commented in the summer 2014 Evaluation Contractor interviews) reported that the pushes had benefited their programs, by helping them focus on high-priority areas and adding a sense of urgency that resulted in more intensive programs for OB-EED, readmissions, and CAUTI.
- **Partnerships added value to HEN strategies.** HENs were asked to list their partnerships and indicate the extent to which each partnership was persistent over time, tightly integrated with their intervention, and brought value to their effort.
  - HENs characterized many of the partnerships as at least moderately integrated with their activities (that is, more than loosely integrated), persisting over time (at least at a medium level), and bringing value to their effort (Figure 3-3).
  - Quality Improvement Organizations (QIOs), health departments, and other HENs were the three most common partners cited by HENs. QIOs and health departments more often were cited as tightly integrated with HEN activities and deemed of critical value (in 9 and 7 HENs, respectively).

> *"We partnered with the Rural Health Coalition. They have been remarkable on the focused attention given to the Partnership for Patients…The group took on no harm across the board from the beginning. Their improvement in no harm is substantial."*–a HEN

Figure 3-3 illustrates the perception of the value of various partnerships in reducing harms.

Figure 3-3—Types of Partner Organizations and Value of Partnerships in Harm Reduction



Source: All 26 HEN-completed spreadsheets describing initiatives and partnerships and rating partnerships in terms of degree of integration, persistence of partner contribution, and value of partnership. Additional information on partnership characteristics is found in Appendix E.
Note 1: Definitions:
Valuable: Partner added value but was less than critical.
Critical Value: Partner contribution was either essential to the effort, or harm reduction would likely have been much less without this partner.
ACOG: The American Congress of Obstetricians and Gynecologists
APIC: Association for Professionals in Infection Control and Epidemiology
SHAs are cited only for the 13 HENs that are not SHAs.
QIOs: Quality Improvement Organizations
Health Departments: State or county
Other: Included but was not limited to: Johns Hopkins Armstrong Institute, other universities, Institute for Healthcare Improvement, Society of Hospital Medicine, the Agency for Healthcare Research and Quality (AHRQ), the Centers for Disease Control and Prevention (CDC), state perinatal collaboratives, state or regional patient safety organizations, state Medicaid agencies, American Board of Internal Medicine (Choosing Wisely campaign), commercial insurers, Institute for Clinical Systems Improvement, California Quality Care Collaborative, Mayo Clinic, other specialty associations, etc.
Note 2: Percentages represent the percentage of HENs listing at least one partnership with the organization type that described the partnership as valuable or critical in at least one adverse event area (AEA) or as a cross-cutting strategy; therefore, a single HEN with 10 partnerships across AEAs might be counted in both the valuable and critical value bars if it described its partnerships with the organization type as being valuable and critical. SHAs participating through AHA/HRET also listed a number of partners that are not presented in this table, because they were not included in the AHA/HRET HEN's spreadsheet.

- **Including the AHA/HRET HEN with a structure that included 31 SHAs as intermediaries to help facilitate local change**. Leadership, networking opportunities, and tools provided by the AHA/HRET HEN allowed subcontracted SHAs to take advantage of the HEN's national infrastructure and effectively target complementary state efforts.

> *"It made so much sense for CMS to invest in hospital associations because the networks were there."* – a SHA subcontractor to the AHA/HRET HEN
>
> *"Participation at local events was greater than AHA/HRET events because hospitals knew the SHA and had a stronger connection to them."*–a SHA subcontractor to the AHA/HRET HEN

As part of PfP's design, HENs had a great deal of flexibility regarding their implementation strategies; they discussed these as most effective:

- Hospital-specific, individualized consultation and technical assistance—whether through virtual or on-site consultation (13 HENs)[3-6]
- Partnership with other organizations—whether newly formed or strengthened (12 HENs)
- Transparency of data reporting and sharing across the HEN's network to promote movement toward open sharing among hospitals (9 HENs)
- Stakeholder engagement, specifically leadership and PFE (9 HENs)
- Presentation of evidence to encourage adoption of best practices (8 HENs)
- Peer-to-peer learning and networking facilitated by the HEN, centered on implementing best practices and addressing barriers (7 HENs)

Participating in any HEN activity greatly increased the probability of a hospital reporting that it made operational changes. For example, across all harm areas and activities, hospitals participating in HEN activities were more likely (often at least twice as likely) to make changes.[3-7] In six harm areas (CAUTI, falls, OB-EED, pressure ulcers, SSI, and readmissions), hospitals participating in value-added networking had the highest probability of making changes due to participation in HEN activities, compared to non-participants, among other HEN activity types. In the remaining areas, hospitals receiving feedback on patient safety data performance from the HEN had the highest probability of making changes due to participation, compared to non-participants, among other HEN activity types. Table 3-4 shows the ratio of relative risk, or probabilities, of a hospital reporting that it made changes due to PfP, for hospitals receiving the HEN activity relative to hospitals not receiving the HEN activity.

> *"We worked through [HEN] on CUSP and TeamSTEPPS [patient safety skills training curricula developed by AHRQ]. They provided a lot last year in education and things we would not have gotten to go to. They had money set aside so quality leaders and staff members could attend. For a smaller hospital money is tight. So being able to go to national meetings is great."*–a visited hospital

[3-6] At least as many HENs as listed for each implementation strategy cited the strategy as being effective.

[3-7] Note: The survey asked about whether hospitals made changes (hospital-wide or unit-specific) due to participation in any HEN activity.

Table 3-4—Relative Risk Ratios for Probability of Making Changes Due to Participation in HEN Activities, by Type and AEA

HEN Activity Type	ADE	CAUTI	CLABSI	Falls	OB-EED	OB-Other	Pressure Ulcers	SSI	VAE	VTE	Readm
Value-Added Networking with Other Hospitals	3.26	1.94	2.84	2.14	2.53	4.30	3.20	3.03	4.09	2.93	2.62
Feedback on Patient Safety Data Performance	3.44	1.88	3.15	2.09	2.25	4.55	2.96	2.85	4.30	2.98	2.50
Skills Training	2.92	1.80	2.35	1.95	2.03	4.02	2.59	2.77	3.53	2.59	2.07
Virtual Coaching and Consultation	2.69	1.71	2.19	1.88	2.02	3.74	2.39	2.66	3.86	2.79	2.19
Other Education and Resources	2.82	1.55	1.94	1.64	2.27	4.13	2.28	2.49	3.30	2.68	2.07
On-site Visits	1.93	1.26	1.61	1.39	1.60	2.88	1.73	1.82	2.23	1.97	1.65

Source: The Evaluation Contractor's Survey of Hospital Participation in Patient Safety Activities, January 2015 through March 2015.
Note: Number of hospitals included ranges from 907 (ventilator-associated events [VAE]) to 2,023 (readmissions). See Appendix C for details on the survey and the simple logistic regression analyses used to calculate the relative risk ratios in the table.
Color Legend:

- Yellow: Hospitals receiving HEN activity type are 1-2 times more likely to have made changes due to PfP in applicable areas than hospitals that did not receive HEN activity type.
- Blue: Hospitals receiving HEN activity type are 2-3 times more likely to have made changes due to PfP in applicable areas than hospitals that did not receive HEN activity type.
- Green: Hospitals receiving HEN activity type are 3 or greater times more likely to have made changes due to PfP in applicable areas than hospitals not receiving HEN activity type.

Factors Related to Specific Harm Area

HENs also identified facilitators applicable to certain focus areas of harm. For example, most HENs (21) reported strong facilitation of their work by national efforts in OB-EED with guidance and support from the March of Dimes and the American Congress of Obstetricians and Gynecologists (ACOG). Consensus around using hard-stop policies (under which hospitals do not schedule OB-EEDs unless there is medical justification) as a key mechanism to reduce OB-EEDs helped to focus HENs' efforts in this area and encourage implementation of OB-EED interventions in their aligned hospitals. In addition, partnerships played a particularly important role in facilitating progress in the areas of ADE, OB-Other, and readmissions. For example, community partnerships and engagement of providers across the continuum of care supported efforts to reduce readmissions.

Barriers

Organizational and External Factors

Hospital electronic health record (EHR) implementation diverted attention from harm reduction work, disrupted data reporting, and was cited as an area for improvement by surveyed hospitals. When aligned hospitals implemented new EHR systems, data reporting was often disrupted and staff attention was temporarily diverted from harm reduction work (20 HENs). Other challenges included difficulty adding new measures and reports to EHRs (7 HENs), lack of data sharing across platforms (5 HENs), and difficulty extracting data to track performance (2 HENs). Several of the 12 visited hospitals also reported temporary patient safety problems resulting from initial EHR implementation. Notably, 4 HENs also reported that established EHRs facilitated change by supporting measurement and standardizing implementation of interventions. When surveyed hospitals were asked to list (open-ended) up to three tools and resources their hospital was lacking that would be helpful in pursuing reductions in adverse events and readmissions, 28 percent of responding hospitals listed improved technology. Examples included an EHR that provides reports, an EHR that is better integrated with physicians and other health care facilities, clinical decision support built into the EMR, and the ability to run reports with less manual data collection.[3-8]

> *"CPOE [computerized provider order entry, within the EMR] generates new errors that you may not have the infrastructure to catch if you don't have vendor support. You're learning as you go. It's new. The efficiency where you're able to bypass multiple steps where there may have been handoff errors can sometimes generate new errors."*–a visited hospital

> *"[The EHR] was definitely a headache at first. But now after we really got familiar with it, it starts to work for you in terms of efficiency and meeting guidelines and it all goes back to better patient care."*–a visited hospital

Hospital consolidation and turnover required "restarts" to regain momentum and continue progress to improve patient safety. Hospital mergers and reorganization of health systems resulted in changes in hospital leadership and staff (16 HENs). Consequently, HENs had to engage new leaders, train new quality improvement staff members, and build new working relationships to resume PfP target harm reduction efforts in hospitals impacted by these transitions.

Resource constraints and competing priorities limited hospitals' capacity to make operational changes to reduce harm. Nearly half of HENs (12) encountered a lack of hospital resources to support work in all areas of harm targeted by PfP and competing priorities, especially in many CAHs. Surveyed hospitals conveyed the same message: when asked to list up to three tools and resources their hospital was lacking that would be helpful in pursuing reductions in adverse events and readmissions, thirty percent of hospitals listed more staff or more time—the most common type of resource that respondents listed.[3-9]

> *"For some of the best practices that we shared, we did not provide the resources to implement. For example, it was a challenge for small hospitals to implement home visits [to prevent readmissions]."*–a HEN

[3-8] *Survey on Prevention of Adverse Events and the Reduction of Readmissions,* spring 2014.
[3-9] *Survey on Prevention of Adverse Events and the Reduction of Readmissions,* spring 2014.

Lack of access to care and inadequate social and financial support increases risk for adverse events, particularly readmissions, among vulnerable patient populations, and meeting these needs requires support outside of the inpatient setting. Patient needs related to various demographic and socioeconomic factors presented challenges for reducing harm (16 HENs). Aging patients and patients with chronic conditions and mental health issues are at a higher risk for harm such as falls and ADE as well as readmissions. Uninsured, Medicaid, and dual eligible patients often lack access to care after discharge and have high rates of noncompliance with care plans, which leads to increased risk of readmission. In addition, poverty, homelessness, and food deserts contribute to poor health among certain patient populations and increase vulnerability to adverse events.[3-10]

> *"After the start of the HEN, we started trending readmissions data…we found that both transportation and medication reconciliation were issues…the one [taxi service] we have doesn't run past 6 PM. Wheelchair transport is $25. County transit stops at 5:30, and you have to book that 24 hours in advance."*–a visited hospital
>
> *"[Mental health] services in this state have been cut consecutively for multiple years, and the burden to manage these patients is totally passed to the hospital. If you're boarding behavioral health patients in the ED because you don't have sufficient beds, that creates a patient safety issue. The ED grinds to a halt. You're redirecting resources to deal with violent patients. There's a finite group of resources, and they can only go so far. ED staff aren't prepared to take care of inpatients, so we have to reallocate staff down there."*–a visited hospital

PfP Campaign-Related Factors

The PfP measurement strategy offered flexibility but lack of standardization and changes during the campaign added burden for HENs and hospitals. PfP offered near-complete flexibility in HENs' and hospitals' selection of outcome measures, then encouraged movement toward common and better measures over time. Many HENs (11 of 16 that commented) perceived this encouragement toward common and better measures as a shift in policy that caused frustration and additional work.

Factors Related to Specific Harm Areas

HENs also identified barriers applicable to certain focus areas. For example, regarding readmissions, barriers to progress reported by 16 HENs included the wide range of evidence-based, best practice prevention strategies in the research literature (with no accepted standard); the lack of standard protocols to ameliorate the impact of patient socioeconomic factors on readmissions; a need for multidisciplinary solutions that go beyond the hospital setting; and cost implications of allocating resources to reduce readmissions while experiencing decreased revenue from readmissions. Barriers related to measurement of harm particularly concerned ADEs, where no standardized measures were available at the start of the campaign, and VAE and CAUTI, where the Centers for Disease Control and Prevention (CDC) definitions changed during the campaign.

> *"Because there are no consensus-defined national measures [for ADEs], we have spent a lot of time on trying to form a consensus on how to best measure this. We changed our measure in January which, looking back, was a bad move. We did it because the hospitals could get to the new data more effectively. On these you lose a lot of time in frustrating discussions trying to collect actionable data and balancing it with the data burden."–a HEN*

[3-10] Food deserts are defined as urban neighborhoods and rural towns without ready access to fresh, healthy, and affordable food.

Hospitals' Engagement with PfP and Implementation of Operational Changes

This section addresses the following research questions related to spread of best practices:

- **Hospitals' engagement and perceptions of PfP**: Did PfP successfully engage hospitals on a national scale? How did participating hospitals perceive PfP?
- **Hospitals' implementation and role of PfP**: Did hospitals make changes to reduce harms and readmissions, including changes to patient care processes, changes to culture and infrastructure for safety, and changes to PFE? To what extent were the changes influenced by participation in PfP as opposed to other factors?
- **Relationship of Processes to Outcomes:** Is there evidence that the types of operational changes made by the hospitals were associated with improved outcomes?
- **Unintended consequences**: What unintended consequences of PfP did HENs and hospitals report?

These questions are answered in the sections below. In sum, the analyses show HENs engaged hospitals widely, that many hospitals made operational changes designed to reduce harm and readmissions, and that a high proportion of these hospitals reported that the changes they made were due to participation in HEN activities. Many hospitals that reported observing improvements in outcomes also reported that the work with their HEN was important to this achievement. Few unintended consequences were reported. Whether these positive self-reported findings translated into measureable changes in outcomes attributable to HENs is discussed in subsequent chapters.

Hospitals' Engagement and Perceptions of PfP

Although HEN outreach was the first step in the improvement process (described in Chapter 1), in order for the HEN work to be effective, hospitals had to actively engage in the HEN activities. Large-scale engagement of hospitals is an essential first step for any initiative seeking to cause national-level change in hospital outcomes. Furthermore, the hospitals must find the activities useful in order for them to continue participating throughout the campaign. This section explains the level of hospital engagement that PfP was able to generate, and it summarizes hospitals' perceptions of PfP. Overall, the Evaluation Contractor finds that hospital engagement was high and hospitals perceived that PfP was useful.

Hospital Engagement

To understand the level of the hospitals' engagement, the Evaluation Contractor used a number of different measures, including whether hospitals were engaged with a HEN, hospitals self-reported participation in HEN-activities, and the extent of the services hospitals received from HENs. The Evaluation Contractor found the following:

Most United States (U.S.) acute care hospitals were aligned with a HEN. PfP achieved widespread engagement, with a total of 3,738 acute care hospitals aligning with a HEN, as of November 2014, representing about 72 percent of hospitals and over 80 percent of admissions to acute care hospitals nationally.

Looking deeper into involvement in PfP beyond alignment with a HEN, more than 65 percent of HEN-aligned hospitals that reported data in CAUTI, OB-EED, falls, CLABSI, SSI, and readmissions described themselves as fully engaged (Table 3-5). This implies that about half of the relevant short-term acute care hospitals nationwide were both participating with HENs and fully engaging in harm reduction efforts in these six key PfP areas.[3-11] Table 3-5 summarizes the HEN-aligned hospitals' report of the degree of engagement with HEN activities by PfP harm area. For consistency with other sections of the report, hospitals that are HEN-aligned for analyses in this chapter are defined based on their presence on HENs' lists of aligned hospitals submitted to CMS in November 2014.

Table 3-5—HEN-Aligned Hospitals' Self-Described Degree of Engagement with HEN Activities During 2012-2013

Harm Area	Fully Engaged	Moderately Engaged	Minimal or No Engagement
CAUTI	69.2%	18.0%	12.8%
OB-EED	68.3%	16.8%	14.9%
Falls	67.9%	17.6%	14.4%
CLABSI	67.2%	17.4%	15.4%
SSI	67.1%	16.8%	16.1%
Readmissions	65.8%	20.7%	13.5%
VAE	56.1%	19.2%	24.7%

[3-11] The average percentage of HEN-aligned hospitals providing the relevant service and reporting full engagement for these six areas is 67.6 percent. Multiplying 0.676 by the 72 percent that were aligned gives 48.7 percent that were both aligned and fully engaged. Hospitals that did not provide the relevant service were excluded from the denominator.

Table 3-5—HEN-Aligned Hospitals' Self-Described Degree of Engagement with HEN Activities During 2012-2013			
Harm Area	**Fully Engaged**	**Moderately Engaged**	**Minimal or No Engagement**
Pressure Ulcers	55.0%	20.1%	24.9%
VTE	53.6%	21.2%	25.1%
All-cause harm	47.1%	23.0%	29.8%
OB-Other	44.5%	25.4%	30.1%
ADE	44.3%	24.0%	31.8%

Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2014.
Note: Hospitals that reported not providing the relevant service were excluded from the denominator for OB-EED, OB-Other, SSI, CLABSI, and VAE, as appropriate. The average number of hospitals included is 822 (range: 784 [OB-Other] to 845 [readmissions]). Estimates shown are weighted to account for sampling and to reduce non-response bias, as explained in Appendix C.

HEN support was deemed not applicable or not needed for some focus areas for many hospitals. While hospitals reported that they were fully engaged in a number of areas (see Table 3-5), typically for other areas either they indicated they were moderately, minimally, or not at all engaged in HEN-sponsored or HEN-led patient safety activities. When hospitals were less than fully engaged in one or more areas, the top three reasons were that the:

1. Area was not applicable to this hospital (42 percent of responding hospitals).[3-12]
2. Hospital already had the improvement support it needed within the hospital or health system (27 percent of responding hospitals).
3. Hospital already sustains zero rates of harm in the area (17 percent of responding hospitals).

Between 6 percent and 11 percent of responding hospitals cited other reasons, including management decision (11 percent), preference to work with another organization (10 percent), lack of resources (9 percent), competing priorities (7 percent), or inconvenient scheduling of learning events (6 percent).

HEN-aligned hospitals in HENs that were owned by healthcare system organizations participated in more types of activities. Examining whether different subgroups of HEN-aligned hospitals reported differences in participation, hospitals in HENs owned by healthcare systems participated in significantly more types of activities on average (out of the six types queried) than hospitals in other types of HENs (Table 3-6) (see Appendix E).[3-13] The six activity types in this analysis were skills training, value-added networking, virtual consultation or coaching, other education or resources, on-site visits, and feedback on patient safety performance data. A summary of the results of the analysis of activity level by subgroup is provided in Table 3-6.

3-12 Note that non-applicability is related to service mix: not all HEN-aligned hospitals provide OB services, ventilator services, perform surgeries, or insert central lines.

3-13 Although the differences for HEN size and rural composition were also statistically significant, hospitals in the more rural HENs and in the smaller HENs participated in less than one more activity type than those in less rural HENs, so this difference was not considered to be meaningful.

Table 3-6—Mean Number of Activity Types (of 6) Hospitals Participated by HEN Subgroup		
HEN or AHA/HRET SHA Characteristic	**Subgroup**	**Mean Number of HEN/SHA Activity Types in Which Hospitals Participated (of these Six: Skills Training, Value-Added Networking, Virtual Consultation, Other Education/Resources, On-Site Visits, Data Feedback)**
Ownership***	System	4.0
	SHA	2.9
	Other[a]	2.3
Size***	< 50 hospitals	3.5
	50-99 hospitals	2.7
	100+ hospitals	2.8
Rural composition**	0-30%	2.6
	> 30%	2.9

Source: The Evaluation Contractor's Survey of Hospital Participation in Patient Safety Activities, January 2015 through March 2015.
Note: N = 2,129 participants; hospitals in HENs with response rates less than 70 percent of network hospitals were excluded from this analysis. The six types of activities were skills training, value-added networking, virtual coaching, on-site visits, performance feedback, and other education and resources. Responses were by PfP focus area, so the means represent the average participation in activities across all relevant areas.
[a]Other HENs include national membership organizations, region-specific and national collaborations, and standing organizations functioning as a resource for quality improvement.
**Statistically significant ($p < 0.05$)
***Statistically significant ($p < 0.01$)

Hospital Perceptions of PfP

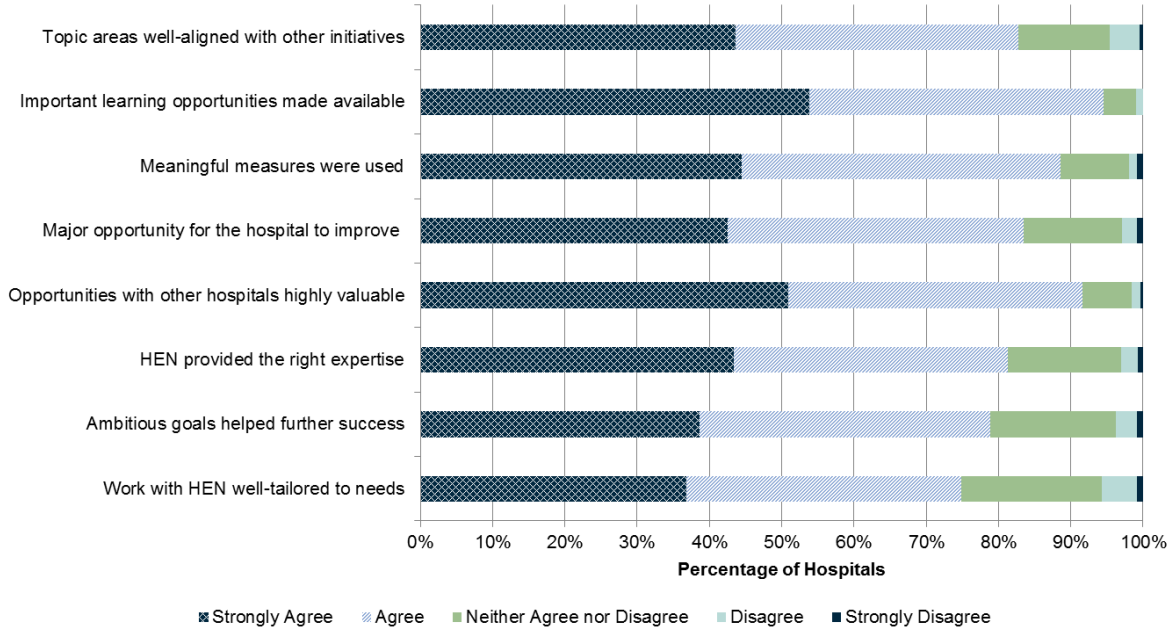
Hospitals Generally Reported a Positive Experience with PfP

- Hospitals found many aspects of the HEN-provided resources helpful to them, and in particular agreed that important learning opportunities had been made available (95 percent), and that opportunities to convene with and learn from other hospitals were valuable (92 percent) (Figure 3-4).
- Over 90 percent reported PfP learning resources were somewhat or very useful in enabling the hospital to take new or different action to reduce adverse events (Figure 3-5).
- Over 800 hospitals (34 percent of the 2,432 respondents to the Survey of Participation in Patient Safety Activities) took the time to write positive comments about PfP in open space at the end of the survey, often complimenting PfP's dissemination of resources and information, HEN staff members or support, or expressing the value of collaboration or networking opportunities (see Appendix E). (Only 5.5 percent of respondents wrote mixed or negative comments.)

> *"With the HEN, there were cutting edge organizations that presented on how they implemented different practices. We could take away things from that that we could maybe implement here."*–a visited hospital

Figure 3-4 illustrates the hospitals' responses to queries regarding how much they valued various HEN activities.

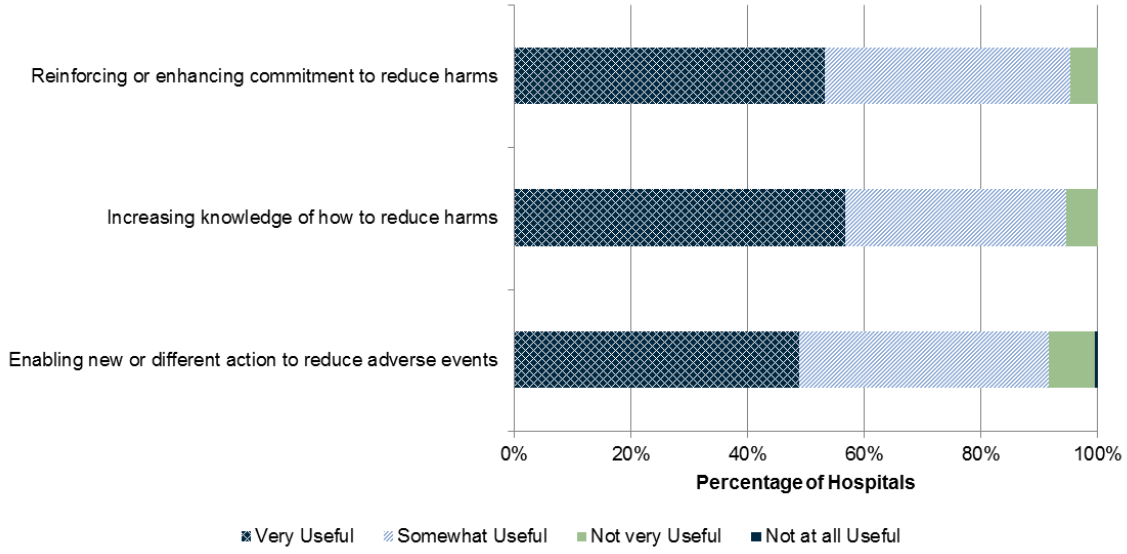
Figure 3-4—Extent to Which Different Aspects of HEN Assistance Were Valued by Hospitals



Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2014.
Note: The average number of hospitals included is 854 (range: 845 [meaningful measures were used] to 862 [major opportunity for the hospital to improve]). Percentages shown are weighted to account for sampling and to reduce non-response bias, as explained in Appendix C.

Figure 3-5 illustrates the hospitals' responses regarding the usefulness of various HEN resources.

Figure 3-5—Usefulness of PfP Resources Accessed by Hospitals



Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2014.
Note: The average number of hospitals included is 857 (range: 855 [enabling new or different action to reduce adverse events] to 859 [increasing knowledge of how to reduce harms]). Percentages shown are weighted to account for sampling and to reduce non-response bias, as explained in Appendix C.

Hospitals with harm reductions report HENs' assistance was important to their achievement. Hospitals self-reported that HENs' assistance was frequently important to harm decreases that they measured during 2012 and 2013. Specifically, of those hospitals that had implemented improvement initiatives and reported observing a 10 percent or more decrease in harm in their data, between 47 and 77 percent cited HEN assistance as important to that reduction, depending on the focus area (Table 3-7, column C). Readmissions stood out, with 77 percent of hospitals that had reported observing a 10 percent or more reduction in their data citing HEN assistance as important.[3-14] The percentages were also high (two-thirds or more hospitals with substantial harm decreases) for CAUTI, falls, OB-EED, and all-cause harm. Column D of Table 3-7 shows the implications of this for the number of estimated hospitals meeting all three criteria—had an improvement initiative in place, saw a 10 percent or greater decrease in harm, and cited the HEN as important to the decrease. That number of hospitals ranged from 441 for VTE to 1,159 for OB-EED.

Table 3-7—Among Those HEN-Aligned Hospitals That Reported a 10 Percent or More Reduction in Harm, Percentage That Cited HEN Assistance as Important to that Reduction

Harm Area	Percentage of HEN-Aligned Hospitals Reporting Improvement Initiative Implemented 2012-2013 (A)	Percentage of HEN-Aligned Hospitals in Column A Reporting Harm Reduction of 10 Percent or More (B)	Percentage of Hospitals in Column B That Cited HEN Assistance As Important to their Harm Reduction (C)	Estimated Number of Hospitals Nationally Meeting Criteria for All Three Columns (3,738 * Column A * Column B * Column C) (D)
ADE	76%	26%	61%	451
CAUTI	88%	40%	70%	921
CLABSI	79%	38%	60%	673
Falls	92%	44%	66%	999
OB-EED	87%	54%	66%	1,159
OB-Other	67%	29%	65%	472
Pressure Ulcers	72%	29%	57%	445
SSI	86%	34%	56%	612
VAE	71%	32%	59%	501
VTE	81%	31%	47%	441
All-Cause	68%	28%	72%	512
Readm	92%	43%	77%	1,139
At Least 1 Area	98%	80%	85%	2,491
At Least 2 Areas	97%	70%	71%	1,802

[3-14] Note that the data that the hospitals were reporting on were not available to the Evaluation Contractor.

Table 3-7—Among Those HEN-Aligned Hospitals That Reported a 10 Percent or More Reduction in Harm, Percentage That Cited HEN Assistance as Important to that Reduction				
Harm Area	**Percentage of HEN-Aligned Hospitals Reporting Improvement Initiative Implemented 2012-2013 (A)**	**Percentage of HEN-Aligned Hospitals in Column A Reporting Harm Reduction of 10 Percent or More (B)**	**Percentage of Hospitals in Column B That Cited HEN Assistance As Important to their Harm Reduction (C)**	**Estimated Number of Hospitals Nationally Meeting Criteria for All Three Columns (3,738 * Column A * Column B * Column C) (D)**
At Least 3 Areas	95%	59%	59%	1,236

Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2014.
[a]The number 3,738 is the number of hospitals HENs reported as aligned as of their last monthly report to CMS, November 2014.
Note: The rows for each harm area include an average of 783 responding hospitals (range is 569 [OB-Other] to 882 [CAUTI]). Estimates shown are weighted to account for sampling and to reduce non-response bias, as explained in Appendix C.

Four of the 12 visited hospitals had some negative comments to share, often along with other positive comments:

- Four did not find very many of the tools and resources that were provided by the HENs very useful.
- One criticized its HEN on two fronts, first for not effectively providing hospital-specific assistance when it was "doing the best [it could], and still floundering," and second for losing "the voice" of the bedside nurses in the HEN activities, which "somewhere along the way tapered off."
- One was disappointed that the HEN's PFE focus with hospitals "was going strong for a year, then it all fell apart," when the key staff member changed positions.

Hospitals' Implementation of Operational Changes and Role of PfP

This section describes the extent to which hospitals reported making the following types of operational changes:

- Changes to patient care processes related to patient safety.
- Changes to better engage patients and families.
- Changes to other aspects of hospital culture and infrastructure that may be important to the hospital's ability to sustain gains made and increase patient safety over time, such as culture of safety, and information system capabilities related to safety.

At the national level, the processes of hospital care and infrastructure that are believed to be important for current and future harm reduction improved during 2012–2013. Hospitals more often attributed changes at least in part to PfP than to factors other than PfP alone.

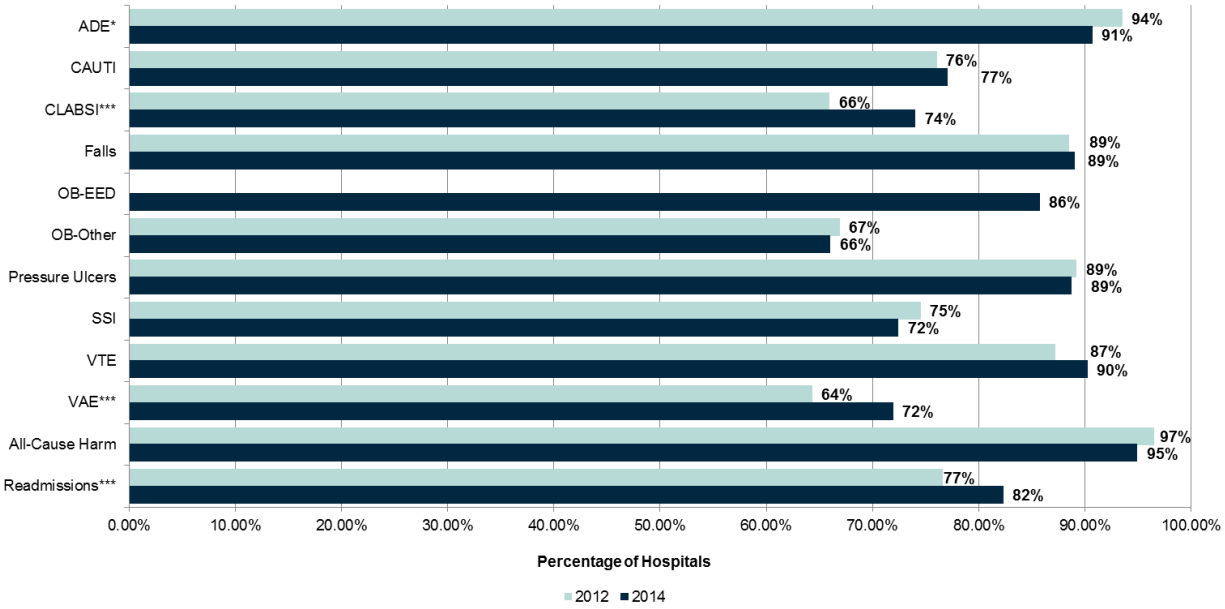
Changes to Patient Care Processes

Most hospitals had begun improvement work on PfP focus areas prior to PfP, and continued these efforts during 2012–2013. A large majority of hospitals (at least two-thirds) reported implementing some type of process improvement effort in each PfP focus area in the 12 months prior to the 2012 survey, and also during 2012–2013 (the period queried in the 2014 survey) (Figure 3-6). The three areas with statistically significant increases in the percentage of hospitals with some effort were: CLABSI, which increased from 66 percent to 74 percent of hospitals with a hospital-wide improvement initiative; VAE, which increased from

64 percent to 72 percent of hospitals with an initiative; and readmissions, which increased from 77 percent to 82 percent of hospitals reporting a hospital-wide initiative. Only one area, ADE, showed a decline (significant, $p < 0.10$), and still over 90 percent of respondents implemented a hospital-wide improvement effort during 2012–2013.

Figure 3-6 demonstrates the percentage of respondent hospitals reporting implementing hospital-wide improvement efforts in 2012 and 2014.

Figure 3-6—Percentage of All Hospitals Implementing Hospital-Wide Improvement Efforts



Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and 2014.

Note: OB-EED was queried in 2014 but not in 2012. For most PfP focus areas, only hospital-wide initiatives were counted, not those focused on specific units; the four exceptions were OB-EED, OB-Other, CLABSI, and VAE, where the bars represent the percentage of hospitals with any improvement effort, not just a hospital-wide effort, since those events are more concentrated within specific units. The percentages in the figure have been rounded for presentation, and are weighted to account for sampling and to reduce non-response bias, as explained in Appendix C. The average number of included hospitals is 577 (range: 300 [OB-EED] to 765 [falls]).

*Change significant, $p < 0.10$

**Change significant, $p < 0.05$

***Change significant, $p < 0.01$

A large number of U.S. acute care hospitals reported making operational changes due to PfP participation, and more so due to PfP relative to other factors. There were 2,101 HEN-aligned hospitals that responded to the Participation in Patient Safety Activities Survey, conducted from January 2015 through March 2015, that reported making hospital-wide or unit-specific changes to care processes in at least one PfP focus area as a result of their participation in HEN activities and receipt of resources, with the average number of areas addressed representing 52.5 percent of applicable harm areas (Table 3-8). Most of these changes were hospital-wide. Factors other than participation in PfP also were reported to drive change, with over 1,300 hospitals reporting that they made changes to at least one area due to factors other than PfP participation (see Chapter 2, Figure 2-37, for a list of typical other factors affecting hospital actions). On

average, hospitals reported addressing 23.5 percent of relevant harm areas with changes based on non-PfP factors, compared with 52.5 percent of relevant areas where changes were made due to PfP.[3-15]

Table 3-8 illustrates the changes in the number of hospitals that reported making operational changes due to PfP or non-PfP factors, and how that declined as the number of harm areas increased.

Table 3-8—Number of Hospitals Documented as Making Operational Changes that They Attribute to Participation in HEN Activities

	Number of Applicable Hospitals	Number of Respondent Hospitals Making Hospital-Wide Changes and Attributing to PfP Participation	Number of Respondent Hospitals Making Either Hospital-Wide or Unit-Specific Changes and Attributing to PfP Participation	Number of Respondent Hospitals Making Changes but Attributing Them to Other Factors (Not to PfP Participation)
Mean Percentage of Applicable Harm Areas[a,b]		45.4%	52.5%	23.5%
At Least One Harm Area	2,344	2,055	2,101	1,346
At Least Two Harm Areas	2,331	1,739	1,852	955
At Least Three Harm Areas	2,312	1,398	1,587	698
Four or More Harm Areas	2,273	1,106	1,311	526

Source: Evaluation Contractor's Survey of Hospital Participation in Patient Safety Activities, January 2015 through March 2015.

[a]For the mean percentage of applicable AEAs (row 1), the following AEAs were considered: ADE, CAUTI, CLABSI, SSI, VAE, VTE, falls, pressure ulcers, OB-EED, OB-Other, and readmissions. For the subsequent rows, safety-across-the-board was also considered.

[b]Applicable focus areas means the hospital provides services that could potentially have a harm in that area. For example, OB-EED and OB-Other harm are only applicable focus areas if the hospital provides OB services. The other harm areas that were inapplicable for some hospitals are SSI (surgical services), CLABSI (provision of central lines), and VAE (ventilator services). (Table 3-6, above, lists all 11 PfP harm areas.)

Survey data identified significantly more implementation of patient safety processes in a majority of specific patient safety processes queried in 2014 compared to 2012.[3-16] Of 44 patient safety practices queried, 26 showed significantly more implementation, including at least one in all ten areas of harm included in the query (Appendix E shows the number of improving practices by area).[3-17] The specific patient safety processes whose implementation increased the most, that is, between 8 and 16 percentage points, were one readmissions process, four OB-EED processes, five CAUTI processes, and one VTE process (Table 3-9). Particularly large gains of 14 to 16 percentage points since 2012 include:

- Readmissions risk assessments: Over one-fourth of hospitals now report providing readmission risk assessment for all patients prior to discharge, up by 16 percentage points.

3-15 Other data confirm process changes at the national level, with 26 of 44 specific care processes showing significant improvement between 2012 and 2014, and identify a subset of these practices (17) where HEN-aligned hospitals newly implemented the practices more than non-aligned hospitals (see Appendix E).

3-16 Survey on Prevention of Adverse Events and Reduction of Readmissions, 2012 and 2014.

3-17 No practices specific to CLABSI were queried, however, between one and eight practices were queried for each of the other 10 harm areas.

- Hard stops for OB-EEDs and related practices: 87 percent of hospitals now report having a hard stop, whereby OB-EEDs are not scheduled, and, related, 88 percent have a protocol for timing or conditions set for OB-EEDs, and 85 percent monitor OB-EED-related protocol monthly, all up 14 to 16 percentage points.
- VTE education to patients and families: Over half the hospitals now report they educate at least 80 percent of their patients on prevention of VTE, up by 15 percentage points.
- Daily review of need for catheter: 85 percent of hospitals now report daily review need for a catheter, up by 14 percentage points.

Table 3-9—Patient Safety Processes Asked of All Hospitals that Improved 8 or More Percentage Points Since 2012

PfP Focus Area: Patient Safety Process	Percentage of Hospitals "Yes" to This Practice in 2014	Percentage Point Improvement 2012-2014
Readmissions: Risk assessment for all patients prior to discharge (n = 1,122)	28%	16%
OB-EED: Protocol for timing/conditions for elective deliveries (n = 697)	88%	16%
OB-EED: Hard stop (no scheduling OB-EED prior to 39 weeks) (n = 640)	87%	16%
OB-EED: Protocol adherence monitored monthly (n = 441)	85%	14%
OB-EED: Program to track/reduce OB-EED (n = 654)	95%	13%
CAUTI: Daily review of need for catheter (n = 1079)	85%	14%
CAUTI: Documentation of reason for catheter insertion (n = 1076)	77%	12%
CAUTI: Auto-stop orders/nurse protocols for catheter removal (n = 1081)	49%	11%
CAUTI: Monitor catheter days (n = 1,083)	92%	8%
CAUTI: Bladder ultrasound scanners (n = 1,046)	50%	8%
VTE: Patients and families education to 80 percent or more patients (n = 1,005)	52%	15%

Source: Hospital Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and spring 2014.
Note: Estimates shown are weighted to account for sampling and to reduce non-response bias, as explained in Appendix C. Number of included hospitals varies by row both because not all hospitals provided obstetrical services (so number of respondents for OB-EED is lower) and because missing responses were excluded from the denominator (no "don't know" response choice was provided, so hospitals likely skipped the question if they did not know the answer).

A smaller subset of patient safety practices (17 of the 44 processes) showed greater implementation among HEN-aligned than non-HEN-aligned hospitals ($p < 0.10$), including at least one process in the areas of readmissions, pressure ulcers, CAUTI, SSI, OB-EED, VTE, and VAE. The largest differences, all with greater than 10 percentage point difference, included:

- Readmission risk assessment for all patients prior to discharge increased from 10 percent of hospitals to 30 percent among HEN-aligned hospitals, whereas it increased from 14 to 21 percent in non-aligned hospitals. Both groups clearly have much room to improve in their consistent and inclusive use of readmissions risk assessment.
- Requiring a check-list-driven protocol for augmentation of labor (such as the Hospital Corporation of America's Pre-Oxytocin checklist) rose from 56 to 63 percent in HEN-aligned hospitals, and dropped from 61 to 52 percent in non-aligned hospitals
- Routine monitoring of the rate at which patients at risk for pressure ulcers receive full pressure ulcer preventive care rose from 66 to 74 percent in HEN-aligned hospitals, while dropping from 67 to 62 percent in non-aligned hospitals
- Requiring a system (such as checklists) to ensure that pressure ulcer risk assessment is conducted within 4 hours of admission for all patients rose from 75 to 78 percent among HEN-aligned hospitals, but dropped from 77 to 69 percent among non-aligned hospitals.

Numerous other organizations and policies beyond PfP also concurrently motivated and provided assistance to hospitals on patient safety improvement. As explained in Chapter 2, hospitals worked on patient safety with organizations other than their HEN, and they reported that a variety of policies and programs beyond HENs had influenced them to take actions to prevent harm and readmissions during the previous 12 months. Hospital system organizations, the CDC's National Healthcare Safety Network (NHSN) reporting system (and presumably associated Medicare requirements to report to it); Medicare payment policies, the Joint Commission's accreditation requirements; and the QIOs are among the top influencers beyond HENs, with between 51 and 64 percent of all hospitals reporting these as influencing their harm reduction actions (see Chapter 2, Figure 2-38, based on the Survey on Prevention of Adverse Events and Reduction of Readmissions, 2014).

> *"Once we see an issue brewing, usually we jump on it…and we augment it with the data from Partnership for Patients, whether it was webinars about infection control, or we've had on-site visits where [HEN] discussed specifically some of the trends and practices going on. So we can take pieces and parts of that, take pieces and parts from other [hospitals in their system] best practices, and the regionalization of our hospital system, and then put our specific plan together."*
> –a visited hospital

Site visits to 12 hospitals confirmed that PfP's role in hospital changes varied, as one would expect in such a complex environment. Among 12 hospitals visited by the Evaluation Contractor, PfP played a critical role in patient safety improvements in two of the hospitals (that is, other factors were clearly less important), and was definitely a contributing factor in four others. In others (4 hospitals), PfP was possibly a contributing factor. In these hospitals, at least some staff members found at least some of the PfP participation to be valuable but were not able to articulate specific contributions to their changes during 2012–2014. In others (2 hospitals), there was no identified role of PfP in the hospitals' changes (one hospital that aligned with a HEN in 2014 and one that never aligned with a HEN). (See Appendix C for site visit methodology and Appendix E for more details about the role of PfP verse other factors in each hospital).

Changes in Culture and Infrastructure for Safety

One of the mechanisms by which PfP intended to reduce harms across a broad set of focus areas was by influencing hospitals' culture and infrastructure for safety (see Chapter 1, Figure 1-1). After a brief explanation of what is meant by "culture of safety," and a note about a technique tried by CMS to generate focus on this topic, this section reviews survey data to identify if changes in culture of safety occurred during the PfP period. In summary, the results below suggest modest progress on many aspects of patient safety culture and infrastructure, with most progress identified for leadership—specifically hospital boards' leadership, the knowledge of what to do to address all PfP focus areas of harm, and the extent to which clinical and non-clinical staff members understand and embrace their role in quality and safety. If the estimated percentage that improved on these dimensions were applied to the number of acute care hospitals nationally, it would imply that in 2014 compared to 2012, over 450 hospital boards increased their leadership role, over 550 more hospitals knew what to do to address all PfP focus areas of harm, about 525 more hospitals reported their clinical staff understood and embraced their quality and safety role, and about 575 more hospitals said this for their non-clinical staff. However, there is clearly further room for improvement as well. For example, even in 2014 only 59 percent of hospitals agreed that the nonclinical staff understood and embraced their role in quality and safety, and only 55 percent reported their information system supports measuring progress on key patient safety indicators associated with the PfP focus areas.

What is meant by "culture of safety" and why should we care about it? Alongside the evolution of measures and care processes related to patient safety, leaders within the healthcare field have highlighted the importance of improving healthcare organizations' culture of safety as a means to reduce harm (Chassin and Loeb 2013; NQF 2009). This focus draws primarily on experience in other high-risk industries such as aviation, where characteristics of safer organizations were carefully studied and where safety improved dramatically after work to improve organizational culture (Reason 1997; Weick and Sutcliffe 2007), although some evidence directly linking hospital culture and safety also exists (Dicuccio 2014; McFadden et al. 2014; Zingg et al. 2015). Key aspects of culture include (Sammer et al. 2009):

- **Leadership**, from the board of directors to the chief executive officer (CEO) to unit managers, who ensure focus is sharp and resources are available to prevent harm throughout the entire organization.
- **Just culture**, which recognizes errors as system failures rather than individual failures and, at the same time, does not shrink from holding individuals accountable for their actions.
- **Learning**, in that the hospital learns from its mistakes and seeks new opportunities for performance improvement. Learning is valued among all staff members, including the medical staff.
- **Evidence-based care**, where patient care processes are based on evidence and processes are designed to achieve high reliability.[3-18]
- **Teamwork and communication**, where spirit of collegiality, collaboration, and cooperation exists; an environment where an individual staff member, no matter what his or her job description, has the right and the responsibility to speak up on behalf of a patient.[3-19]
- **Patient-centered care**, where care is centered on the patient and family (discussed in the next section).

[3-18] This component was addressed partially through the analyses of patient safety processes described above. Although the evidence-based care aspect of culture was not viewed as appropriate for a simple query on the survey, the survey did ask about 44 specific patient care processes believed related to safety. Systematic evidence review was not conducted during survey development due to time constraints, so a hospital's use of the 44 processes is not considered a precise indicator of the extent to which it uses evidence-based practice.

[3-19] The teamwork and communication component was not able to be addressed with the methods available. Assessing teamwork and communication within a hospital would be difficult and more intrusive than would be tolerated in a large-scale evaluation, requiring methods such as conducting a series of observations at each hospital.

CMS developed first-generation process metrics for leadership and PFE as a way to increase hospital focus on these aspects of culture. Nationally standardized, tested, endorsed metrics do not exist to measure leadership and PFE, despite their recognition as key aspects of patient safety culture. Therefore, to increase hospitals' focus on leadership and PFE, CMS worked with stakeholders in the learning community to create a consensus-based set of process metrics covering four leadership processes and five PFE processes (Table 3-10). HENs asked their network hospitals if they had implemented each of these, and periodically re-assessed their implementation. The extent to which hospitals changed from "no" to "yes" for implementing these processes was incorporated into HENs' interim assessments for 2014. The collection of information about implementation of these first-generation metrics led to a great deal of discussion about these topics between HENs and their aligned hospitals. The need for greater specificity to facilitate common interpretation of process metrics, and the need for analysis to tie leadership and PFE processes to outcomes became clear as a result of this experience.

Table 3-10—PfP's First-Generation Leadership and PFE Metrics

Leadership Metrics	
1	Hospital has regular quality review aligned with the PfP goals.
2	Hospital has a public commitment to safety improvement with transparency in sharing more than CORE measurement data with the public.
3	Hospital staff, all or nearly all, have a role or perceived goal in patient safety (e.g., Can be explicit in human resource [HR] goals or a group bonus based on a patient safety target).
4	Hospital board of trustees has a quality committee established; with regular review of patient safety data, including review and analysis of risk events.
PFE Metrics	
1	Prior to admission, hospital staff provides and discusses with every patient that has a scheduled admission, allowing questions or comments from the patient or family, a planning checklist that is similar to CMS' Discharge Planning Checklist.
2	Hospitals conduct shift change huddles and do bedside reporting with patients and family members in all feasible cases.
3	Hospital has a dedicated person or functional area that is proactively responsible for patient and family engagement and systematically evaluates patient and family engagement activities.
4	Hospital has and active PFE Committee OR at least one former patient that serves on a patient safety or quality improvement committee or team.
5	Hospital has at least one or more patient(s) who serve on a governing or leadership board and serves as a patient representative.

Source: For complete discussion of the Z-5 scoring criteria, see PfP PEC: Hospital List Scoring Criteria and HEN-Wide Performance Benchmarks, April 2014. http://www.healthcarecommunities.org/SearchResults/ViewDocument.aspx?EntryId=78514&CategoryID=45041.

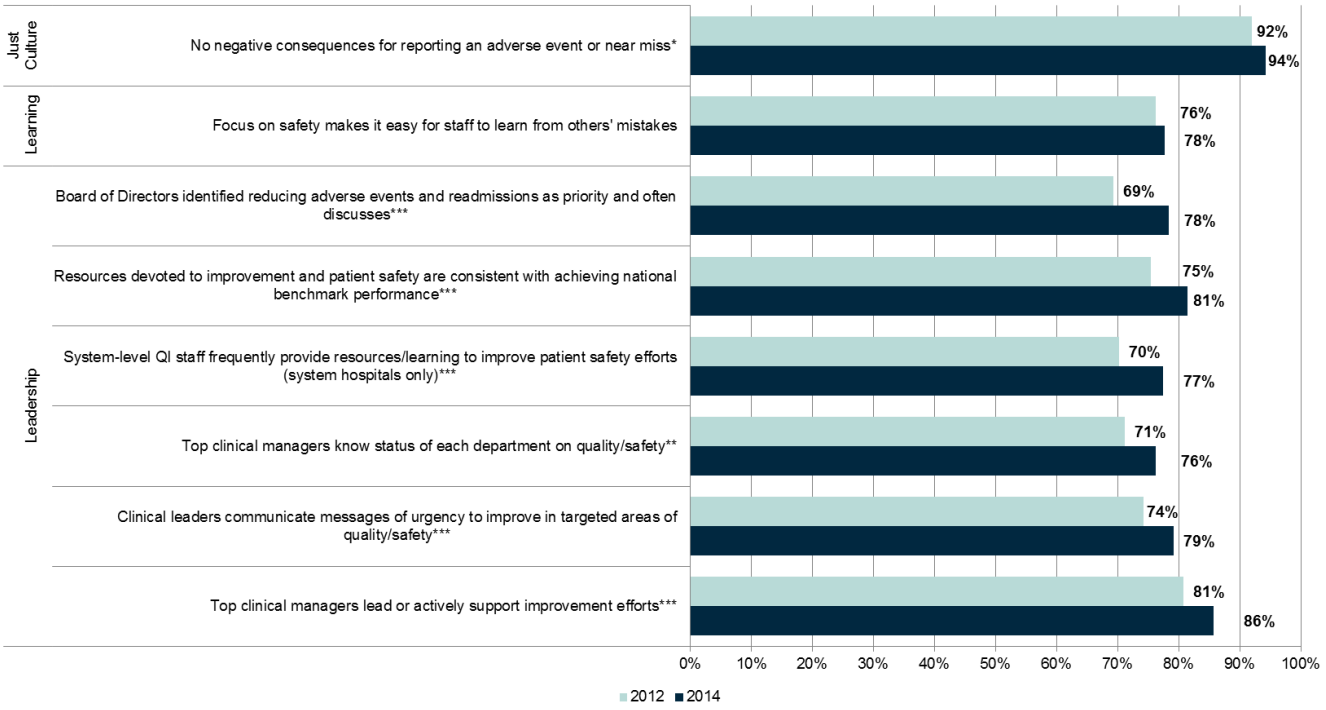
Leadership improved (Figure 3-7). Every aspect of hospital leadership's role in patient safety that was queried improved significantly (although not always to a large degree), from board-level prioritization of patient safety and discussion of safety topics, to top clinical managers knowing the status of every department with respect to patient safety and quality metrics. The largest change was that hospital boards are now more often reported to have identified reducing adverse events as a priority and discuss this often (69 percent in 2012, 78 percent in 2014).

"Our whole culture on the management and medical staff has definitely changed since 2012. Our medical staff meetings are totally different than what they used to be. Quality was a small piece of the medical staff meetings. Now quality is a good portion of our meeting."–a visited hospital

Just culture and learning did not progress, or not by much, although the just culture results were the highest among the aspects studied (Figure 3-7). By 2012, most hospitals agreed with the statement that "reporting an adverse event or a near miss will not result in negative repercussions for the person reporting it." The percentage increased slightly in 2014 from 92 to 94 percent. There was no significant increase in hospitals' agreement with the statement: "The hospital's focus on safety makes it easy for staff to learn from others' mistakes;" 76 to 78 percent of hospitals agreed in both periods.

> *"As we've ramped up our program [to get staff to report near misses], we've seen a significant increase in occurrence reporting (we expect that, that's what we want), but also a significant drop in Serious Safety Events."*–a visited hospital

Figure 3-7—Changes in Hospital Leadership, Just Culture, and Learning



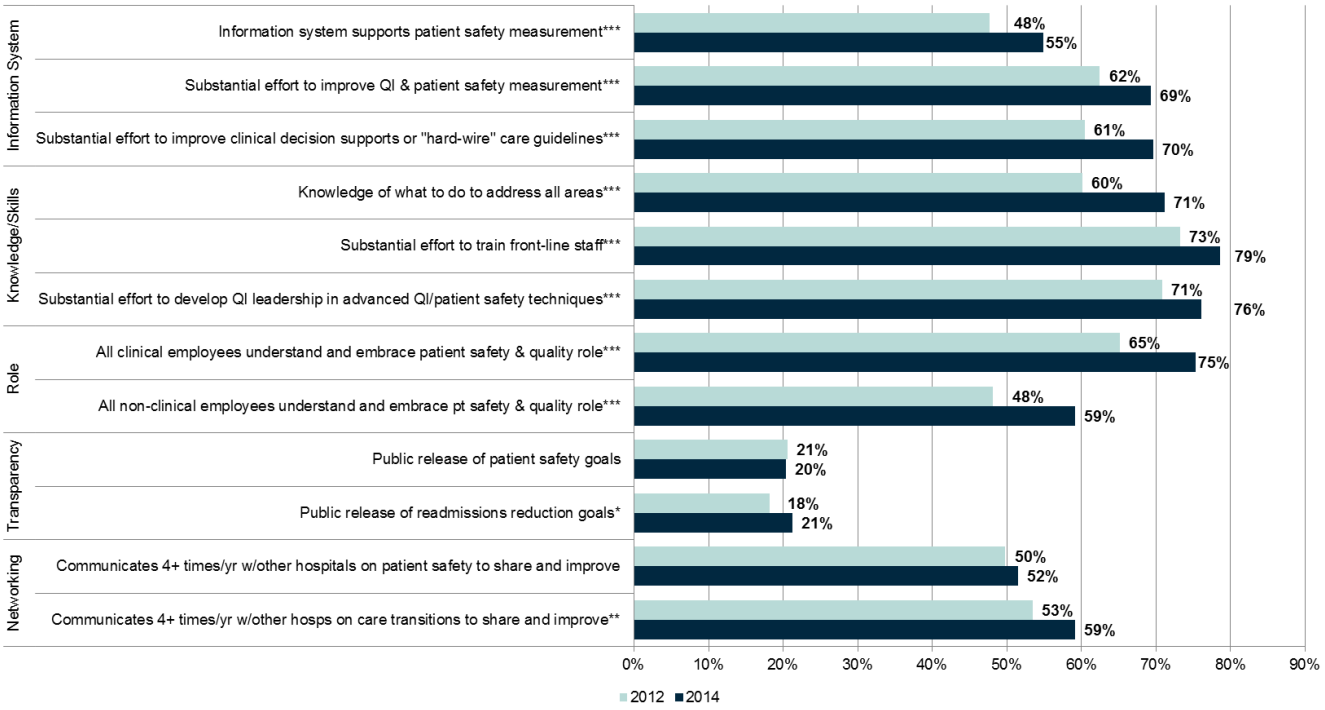
Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and 2014.
Note: The average number of included hospitals is 1,113 for these items (range 1,106 to 1,118) except for one item that was restricted to system hospitals (n = 557) (system level QI staff frequently provide resources/learning to improve patient safety efforts). Estimates shown are weighted to account for sampling and to reduce non-response bias, as explained in Appendix C.
*Significant change, $p < 0.10$.
**Significant change, $p < 0.05$.
***Significant change, $p < 0.01$.

Beyond these traditionally defined components of culture, the analyses below provide potentially important supplementary information about other aspects of hospital infrastructure and culture that are related to hospitals' ability or motivation to prevent harm (Figure 3-8):

- **Information system capability**: The capability of the hospital's information system to measure and provide performance feedback on harm reduction to support change and sustainability (Greenhalgh et al. 2004; Perla et al. 2013; Yuan et al. 2010). Hospital information systems improved in their ability to measure harm, and more effort is being made in 2014 relative to 2012 to improve clinical decision supports or "hard-wire" care guidelines, as well as to improve measurement of quality and patient safety, according to respondents.
- **Knowledge/skills**: Hospitals' self-reported level of skills and knowledge needed to make changes to reduce patient harm (Greenhalgh et al. 2004; Damschroder et al. 2009). Hospitals' perceived sufficiency of knowledge on how to address harm in all PfP focus areas improved from 60 percent "yes" to 71 percent "yes" between 2012 and 2014. The percentage of respondents agreeing their hospital made "substantial" effort to train staff members and leadership in patient safety improvement improved as well.
- **Role**: Engagement of clinical and non-clinical hospital staffs in understanding and embracing their role in quality and safety (Greenhalgh et al. 2004; Damschroder et al. 2009). Many more hospitals reported their clinical and nonclinical staffs understand and embrace their roles in quality and safety in 2014 relative to 2012. There remains substantial opportunity to further improve on this aspect of culture; however, particularly for nonclinical staff members, where even with the improvement, only 59 percent of hospitals in 2014 agreed nonclinical staff members understood and embraced their role.
- **Transparency of goals**: (In this case, patient safety and readmissions reduction goals) and alignment of performance feedback with goals (Damschroder et al. 2009). Hospitals made slight progress in their willingness to publicly release readmissions goals, but there was no change in willingness to publicly release patient safety goals. Most hospitals indicated they still did not do this (only 20 and 21 percent did, respectively).
- **Networking** (Damschroder et al. 2009): Hospitals were slightly more inclined to network (as measured by communicating with other hospitals at least 4 times/year) regarding their care transitions initiatives than in 2012, but did not increase networking (by this definition) regarding their patient safety initiatives.

Figure 3-8 illustrates the changes in responses regarding several aspects of hospital culture of safety between 2012 and 2014.

Figure 3-8—Changes in Several Aspects of Hospital Culture and Infrastructure to Support Patient Safety



Source: Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and 2014.
Note: The average number of included hospitals is 1,089 (range: 1,013 (public release of patient safety goals) to 1,126 (substantial effort to develop QI leadership in advanced QI/patient safety techniques). Estimates shown are weighted to account for sampling and to reduce non-response bias, as explained in Appendix C.
*Significant change, $p < 0.10$.
**Significant change, $p < 0.05$.
***Significant change, $p < 0.01$.

For most aspects of patient safety culture, HEN-aligned hospitals showed a similar pattern of change to non-aligned hospitals.[3-20] The two exceptions were:

- *Communication with other hospitals to improve safety and reduce readmissions*: HEN-aligned hospitals, more than others, increased networking with other hospitals about patient safety and care transitions for purposes of information sharing and improving each other's initiatives. Specifically, 53 percent of HEN-aligned hospitals reported communicating with other hospitals about patient safety initiatives 4 or more times per year in the 2012 survey, rising to 58 percent in 2014, whereas just 43 percent of hospitals did so in 2012, and this decreased to 36 percent by 2014 (difference significant, $p < 0.05$) (data not shown). The pattern was even more striking for communications about care transitions initiatives, with 53 percent rising to 63 percent communicating with other hospitals at least four times per year, compared to 55 percent dropping to 50 percent among other hospitals (difference significant, $p < 0.01$).
- *Public release of patient safety reduction goals*: more HEN-aligned hospitals than others began or continued publicly releasing improvement goals for patient safety between 2012 and 2014. Specifically, 22 percent of HEN-aligned hospitals publicly released patient safety goals in 2014, up from 20 percent in 2012, whereas only 15 percent of other hospitals did so, and that was a drop from 22 percent in 2012 (difference significant, $p < 0.10$).

Changes in PFE[3-21]

Although understanding how to measure PFE and its potential for contributing to reductions in harm and readmissions is still in its infancy, patients are at the center of the care process and unquestionably bring unique knowledge of their condition and care experience, as well as a continual point of observation within the complexity of the hospital environment. The Evaluation Contractor was able to analyze two aspects of PFE to the extent to which hospitals reported: (1) including patients and familiesin structures that address patient safety, such as the board and patient safety committees, and (2) involving patients and families more at the point of care.

Some progress was found on both aspects, with more progress involving patients and families at the point of care than with including them in hospital structures that address patient safety. On 6 of 10 measures that could be compared, HEN-aligned hospitals improved significantly more than non-aligned hospitals ($p < 0.10$).[3-22]

In addition to reporting improvement on specific PFE processes, 21 percent of HEN-aligned hospitals and 15 percent of non-aligned hospitals reported that compared to two years ago, there was improvement in the extent to which patients' and families' perspectives are heard and considered in ways important to the safety of patient care (difference significant, $p < 0.10$).

Integration of Patients into Patient Safety-Related Structures

Summary of results presented below. Hospitals integrated patients more into advisory and patient safety improvement committees and many assigned a leader for PFE during 2012–2014. However, progress in

3-20 Based on the Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and 2014. Hospitals' alignment with HENs were defined in this chapter as hospitals on HENs' lists of aligned hospitals as of June 2012, for consistency with other parts of the report.

3-21 The information in this section is based on the Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and 2014, except where otherwise noted.

3-22 Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and 2014.

integrating them into boards is less clear, and integration into root cause analysis and partnership for monitoring compliance with patient safety practices actually declined.

Patients began to be more often included on PFE Advisory or Improvement Committees. Hospital survey data indicate a significant but modest increase in the percentage of hospitals including patients on a patient and family advisory committee (growing from 26 percent to 31 percent) and on patient safety committees (growing from 10 to 14 percent). HEN-aligned hospitals improved more than others, on both fronts.[3-23]

- *Advisory councils*: Among HEN-aligned hospitals, the percentage of hospitals where patients actively take part in advisory councils grew from 29 to 36 percent, whereas this remained about the same (19 percent and 18 percent in the 2 years, respectively) for the non-aligned group (difference significant, $p < 0.05$).
- *Patient safety committees*: Among HEN-aligned hospitals, the percentage of hospitals with patients participating on patient safety committees grew from 11 percent to 16 percent, whereas this stayed about the same in non-aligned hospitals (7 percent and 8 percent in the 2 years, respectively) (difference significant, $p < 0.10$).

Many hospitals newly established a designated leader for PFE. Over 980 HEN-aligned hospitals newly assigned a leader for PFE (data for HEN-aligned hospitals only) (Table 3-11, based on HEN Z-5 tables).

Patients' representation on hospital boards may have grown. Nine hundred HEN-aligned hospitals may have added patient representation to their hospital boards, according to HEN-reported data, although the survey data do not find progress on this aspect of PFE. In addition, the survey results showed a change in the unexpected direction for HEN-aligned hospitals, dropping slightly among HEN-aligned hospitals (53 to 50 percent), while increasing in non-aligned hospitals (36 to 43 percent) (difference significant, $p < 0.05$).[3-24]

Patients' involvement in root cause analysis and as partners in monitoring compliance with patient safety practices decreased slightly.[3-25] Hospitals in 2014 less often reported patients participating in root cause analysis and acting as partners in monitoring compliance with patient safety practices, although the decline for root cause analysis was less among HEN-aligned than other hospitals ($p < 0.10$). In 2014, only 4 percent of hospitals nationally included patients in root cause analysis, and only 23 percent of hospitals included patients as partners in monitoring compliance with patient safety practices.

Table 3-11 presents the implementation of formalized integration of patients into safety culture structures from 2013 to 2014.

3-23 The hospital survey estimates imply that approximately 250 more HEN-aligned hospitals included patients on an advisory council and about 210 began including patients on patient safety committees. However, the HEN-reported data (HEN Z-5 tables) suggest much greater improvement, with over 900 hospitals newly indicated to be including patients on a PFE or patient safety improvement committee (Table 3-7). Typically, the hospital survey data would be viewed as the more credible source, since the hospitals respond directly rather than through the HEN. However, it is possible the single hospital respondent to the survey was not aware of the full extent of patient participation, whereas the HEN may have taken more care to identify other relevant contacts so as to fully characterize the extent of patients' participation.

3-24 The difference between the HENs' and hospitals' reports of change in patients' participation on the board (that is, between the HEN Z-5 tables and responses to the Survey on Prevention of Adverse Events and Reduction of Readmissions) may be due to differing interpretation of the question by the different stakeholders. Although the percentages of hospitals with a patient on the board are not very different in 2014 between the hospitals' reports and the HENs' reports (48 percent vs. 42 percent), it may be that HENs were initially more conservative in interpreting the metric, and so seem to show greater growth. The issue arises as to how to count community leaders who serve as board members for small, rural hospitals; for example, where the individuals were chosen because of their community leadership but where they are also patients from time to time.

3-25 Survey on Prevention of Adverse Events and Reduction of Readmissions, spring 2012 and 2014.

Table 3-11—Change in HEN-Aligned Hospitals' Integration of Patients into Patient Safety-Related Structures			
	Number of HEN-Aligned Hospitals Reported as Not Having This Process (Since July 2013) (A)	Number of HEN-Aligned Hospitals Newly Implementing This Process (Through November 2014) (B)	Percentage of Column A Hospitals Newly Implementing This PFE Process (C)
PFE leader assigned: Hospital has a person (or functional area) who may also operate within other roles in the hospital, who is dedicated and proactively responsible for PFE and systematically evaluates PFE activities (e.g., open chart policy, PFE trainings, establishment and dissemination).	1,640	989	60%
Patient on PFE committee or improvement committee: Hospital has an active PFE committee OR at least one former patient who serves on a patient safety or quality improvement committee or team.	1,375	914	66%
Patient on board: Hospital has at least one or more patient(s) who serve on a governing and/or leadership board and serves as a patient representative.	1,227	911	74%

Source: HEN Z-5 tables submitted monthly, July 2013 through November 2014, which provided the HEN's report of each hospital's status on each metric (yes, no, or unknown).
Note: The metrics listed in the table were established by CMS in February 2013 by consensus, after consultation with knowledgeable stakeholders.

PFE Involvement at the Point of Care

Many hospitals began discharge planning with the patient prior to admission. Over 1,000 HEN-aligned hospitals newly implemented a process to work with all patients to use a checklist to begin discharge planning even prior to admission (data for HEN-aligned hospitals only) (Table 3-12, based on HEN Z-5 tables).

Many hospitals began to involve patients and families more often in "shift change huddles," where the nurse or care team discusses the patient's condition and plan of care. Over 900 HEN-aligned hospitals were reported to shift from not doing this in all feasible cases, to doing it in all feasible cases (data for HEN-aligned hospitals only). However, the survey data indicate a more modest increase in the percentage of hospitals that said that patients were invited to participate in multidisciplinary rounds (from 21 to 24 percent, significant $p < 0.10$).[3-26]

"When we first started [bedside reporting] it wasn't easily accepted because they [nurses] weren't used to giving info in front of the patient and family. It was a big HIPAA thing was that they didn't want to talk in front of the family. They were doing bedside reporting, then doing a different report at the nursing station. But it has gotten a lot better…they ask the patient if it's OK to talk in front of their family, and if they say no they'll ask them to step out."
–a visited hospital

Education to patients and families on pressure ulcers and VTE increased. Among HEN-aligned hospitals, the percentage of hospitals that report educating at least 80 percent of their patients on the risk of VTE and its prevention rose from 37 to 53 percent, whereas in the non-aligned group this rose from 41 to 49 percent (difference significant, $p < 0.05$). Both HEN-aligned and non-aligned hospitals increased routine education on pressure ulcers; the overall percentage rose from 51 percent in 2012 to 58 percent in 2014.

Hourly rounds to check on patient comfort needs increased. The percentage of hospitals reporting they conduct hourly comfort rounding to address patient needs for pain relief, toileting, and positioning increased from 70 to 76 percent.

Table 3-12 illustrates the change in PFE involvement at the point of care among HEN-aligned hospitals from 2013 to 2014.

Table 3-12—Change in HEN-Aligned Hospitals' Involvement of Patients and Families at the Point of Care

	Number of HEN-Aligned Hospitals Reported as Ever Without This Process (Since July 2013) (A)	Number of HEN-Aligned Hospitals Newly Implementing This Process (Through November 2014) (B)	Percentage of Column A Hospitals Newly Implementing This PFE Process (C)
Discharge planning checklist: Prior to admission, the hospital staff provides and discusses a planning checklist with every patient who has a scheduled admission—allowing for questions or comments from the patient or family—a planning check list that is similar to CMS' Discharge Planning Checklist.	1,640	1,091	67%

[3-26] The implication given the different wording of the metric in the HEN data versus the survey may be that there is much more frequent involvement of the patient in routine nursing huddles, but perhaps this is not extending as often to increase routine involvement where the physician and broader care team are present.

Table 3-12—Change in HEN-Aligned Hospitals' Involvement of Patients and Families at the Point of Care			
	Number of HEN-Aligned Hospitals Reported as Ever Without This Process (Since July 2013) (A)	Number of HEN-Aligned Hospitals Newly Implementing This Process (Through November 2014) (B)	Percentage of Column A Hospitals Newly Implementing This PFE Process (C)
Shift change huddles/bedside reporting: Hospital conducts shift change huddles and does bedside reporting with patients and family members in all feasible cases.	2,322	900	39%

Source: HEN Z-5 tables submitted monthly, July 2013 through November 2014, which provided the HEN's report of each hospital's status on each metric (yes, no, or unknown).
Note: The metrics listed in the table were established by CMS in February 2013 by consensus, after consultation with knowledgeable stakeholders.

Association Between Patient Safety-Related Processes and Outcomes

Many PfP initiatives aimed to educate hospitals on processes of care that could reduce the risk of adverse events. Therefore, the PfP campaign may have resulted in changes in hospital processes that are related to improvements in outcomes. In this analysis, the Evaluation Contractor examined whether hospitals that implemented during our study period eight pre-identified processes of care across three areas of harm (readmissions, pressure ulcers, and VTE) had better improvements in outcomes on those adverse events than other hospitals.

Specifically, the Evaluation Contractor compared mean change in the 2011 and 2013 outcome rates among hospitals that implemented the processes of care between 2012 and 2014 to mean change among those that did not.[3-27] To measure processes of care, information on the 2012 and 2014 rounds of the Survey on the Prevention of Adverse Events and Readmissions was used. The Evaluation Contractor defined hospitals as newly implementing a process if they reported not having the process in place in the first round of the survey, but having it in place in the second round.

The outcomes examined in this analysis are measured based on Medicare fee-for-service claims data, and include the Medicare fee-for-service 30-day all-cause readmissions measure and the AHRQ patient safety indicators (PSIs) for pressure ulcers (PrU [PSI-03]) and post-operative (VTE [PSI-12]).[3-28]

The number of surveyed hospitals that implemented any specific process between the two rounds of the survey was small, ranging from just 26 hospitals that implemented a standard risk assessment tool for pressure ulcers to 230 that implemented a counseling approach that provides information to patients on VTE risk and prevention. With this level of new implementation among the hospitals sampled for the survey, it makes it difficult to find statistically significant results; and indeed the Evaluation Contractor only found one – hospitals that began routine education to families about pressure ulcer prevention and documentation of the education in the chart had a significant decrease in pressure ulcer rates. That said, the Evaluation Contractor did find that the remaining estimates were all in the expected direction, and many would be clinically meaningful if accurate. That is, the hospitals that implemented a specific process improved more in that area than hospitals that did not implement the process. This is suggestive that there is a relationship between these

3-27 This analysis includes all surveyed hospitals, both HEN-aligned and non-HEN-aligned.
3-28 The definitions for the AHRQ PSI indicators can be found here: http://www.qualityindicators.ahrq.gov/modules/PSI_TechSpec.aspx.

processes and relevant outcome measures, although the Evaluation Contractor cannot rule out that it is due to chance.

Table 3-13—Relationship between Hospital Processes of Care and Change in Medicare Adverse Event Rates

	Hospitals that Implemented Process	All Other Hospitals
Readmissions[a]		
Standardized medication reconciliation at time of discharge or transfer for all patients		
Mean change in outcome rate 2011-2013	-1.6	-1.3
Mean change in outcome rate (just above) as a percentage of mean outcome rate for 2011	-10.2%	-7.7%
N	38	887
Completion of risk assessment for all patients, prior to discharge		
Mean change in outcome rate 2011-2013	-1.4	-1.3
Mean change in outcome rate (just above) as a percentage of mean outcome rate for 2011	-8.0%	-7.7%
N	196	542
Pressure Ulcers[b]		
System for completion of risk assessment within 4 hours of admission, for all patients		
Mean change in outcome rate 2011-2013	-0.030	0.004
Mean change in outcome rate (just above) as a percentage of mean outcome rate for 2011	-19.0%	2.9%
N	119	777
Use of standard risk assessment tool		
Mean change in outcome rate 2011-2013	-0.030	-0.003
Mean change in outcome rate (just above) as a percentage of mean outcome rate for 2011	-21.7%	-1.9%
N	26	895
Routine education to patients and families and documentation of this in the chart		
Mean change in outcome rate 2011-2013	-0.075	0.022*
Mean change in outcome rate (just above) as a percentage of mean outcome rate for 2011	-42.7%	17.3%
N	187	694
Monitor rates of full preventative care among patients at risk for pressure ulcers		
Mean change in outcome rate 2011-2013	-0.021	0.002
Mean change in outcome rate (just above) as a percentage of mean outcome rate for 2011	-12.5%	1.8%

Table 3-13—Relationship between Hospital Processes of Care and Change in Medicare Adverse Event Rates		
	Hospitals that Implemented Process	**All Other Hospitals**
N	169	721
VTE[b]		
Hospital wide, written prophylaxis policy for VTE		
Mean change in outcome rate 2011-2013	-0.070	0.011
Mean change in outcome rate (just above) as a percentage of mean outcome rate for 2011	-36.8%	5.7%
N	116	782
Provides information on risk of VTE and its prevention to 80 percent or more patients		
Mean change in outcome rate 2011-2013	-0.022	0.013
Mean change in outcome rate (just above) as a percentage of mean outcome rate for 2011	-10.5%	7.5%
N	230	610

Source: Evaluation Contractor's analysis of the *Survey on the Prevention of Adverse Events and Readmissions* and Medicare inpatient claims data.

Note: This table shows the mean change in the outcome rates from 2011 to 2013 among Medicare patients. The number of hospitals used for each comparison is also displayed. The number of observations are different for each item due to item non-response in the survey. Estimates shown are weighted to account for sampling and to reduce non-response bias, as explained in Appendix C.

[*]Means for the two groups are significantly different at the 0.05 level, two-tailed test.

[a]30-day all-cause readmissions per 100 discharges

[b]Reported as the number of adverse events per 1,000 Medicare beneficiary discharges at risk

Unintended Consequences

While all of the PfP campaign's policies and initiatives were designed to generate improvements in patient safety, sometimes the stakeholders responded in unanticipated ways that thwarted progress or created other issues that need to be considered when designing or considering future programs.[3-29]

> *"Implementing 11 areas at one time, we received a lot of pushback about reporting burden. In addition to the 15 or 16 quality initiatives that [private insurer] runs separate from us, any given hospital is doing 30 or 40 quality initiatives. That's probably the biggest thing we've heard…our concern is the burden of the effort"*–a HEN

Only 7 of 25 HENs and 11 of 24 SHAs working with the AHA/HRET HEN identified any negative unintended consequences.[3-30] The most-often cited responses fell into two categories of unintended consequences:

1. **Increased hospital workload (4 HENs, 10 SHAs)**. HENs and SHAs described the intensity of hospital work due to PfP and concurrent quality improvement initiatives; in a few cases, HENs/SHAs also attributed workload and "data burden" on hospitals as cause for individual hospitals exiting the campaign. In addition, the issue was raised by non-federal partners (4 of 20) and visited hospitals (5 of 10).
2. **Fragmentation of established SHA networks (4 HENs, 1 SHA)**. One SHA (either HENs that were also SHAs or SHAs participating through the AHA/HRET HEN) in particular described fragmentation or dilution of established networks. One non-federal partner interviewee also commented on this, viewing PfP as "disruptive" to one of their initiatives in place in a state, describing a state collaborative of hospitals that was "splintered" by the participating hospitals joining different HENs.

> *"We supported everyone in joining the HEN of their choice, but through this we lost the engagement of some of our hospitals… If we collected one measure and concentrated on something different than other HENs, it diluted the hospitals' commitment to working with us on statewide initiatives."*–a HEN

In addition, two HENs described hospital revenue loss from reduced admissions as a consequence of readmissions reductions when asked about unintended consequences. As a follow-up, the Evaluation Contractor examined the readmissions discussions in the fall interview notes to see if any more instances of this issue were raised; in total, 4 HENs and 8 SHAs described economic challenges associated with readmissions reduction, particularly for critical access and small hospitals, although not as an unintended consequence. One HEN and one SHA also described challenges with hospital engagement due to measurement changes during the campaign.

[3-29] This analysis is based primarily on interviews conducted with 26 HENs and 24 AHA/HRET SHAs in fall 2014, during which participants were asked about unintended consequences—both positive and negative—that they thought hospitals participating in PfP might have experienced. Interviews with non-federal partner organizations and site visits to hospitals were also used to corroborate or expand on the HENs' statements.

[3-30] The Evaluation Contractor did not ask one HEN, which participated in a pilot interview, about unintended consequences.

4. Quantitative Analysis of the Overall HEN Component Impact on Observed Outcome

This chapter presents a variety of quantitative analyses designed to assess the Partnership for Patients (PfP) campaign. The first section examines the changes in the Hospital Engagement Network (HEN)-level data trends over the time of the PfP campaign; the second section presents the results of analyses comparing outcomes between the HEN-aligned hospitals and a comparison group of non-HEN-aligned hospitals using national data.

Analysis of HEN-Level Data

Over the course of their participation in PfP, the HENs collected data on a wide variety of measures in each topic area. In contrast to the national leading indicator measures that were examined in the first part of this chapter, the HEN-level data allow for examination at a local level. Rather than assessing the change in a national rate, the HEN-level data provide the detail to understand which HENs achieved success, and in which topic areas. Additionally, the increased granularity of the HEN-level measures allows assessment of the variability in results across HENs. This section of the report presents an assessment of HEN-level data, stratified by AEA, to better understand which patient harms were reduced in more HENs, by greater amounts, and whether or not the aggregated trends across measures provide evidence that would be consistent with quality improvement changes occurring during the time frame of the PfP campaign. A detailed discussion of the creation of the HEN-level data is included in Appendix D.

Interrupted Time Series (ITS) by Adverse Event Area (AEA)

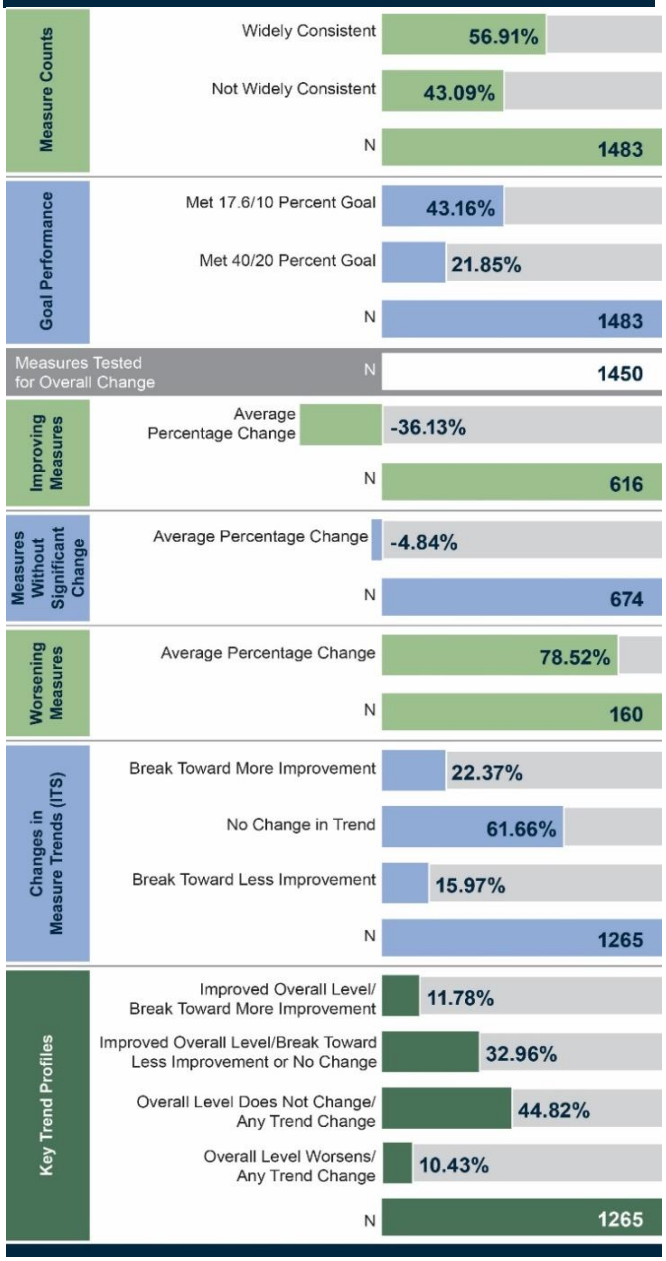
Several different analyses were performed. Each outcome measure was evaluated for whether it met a threshold to be deemed widely and consistently reported among HEN hospitals. The extent of improvement in terms of relative (percentage) change and whether or not the measure achieved goal was examined next. The Evaluation Contractor assessed outcomes with sufficient observations using ITS analysis to detect the presence and direction of any structural breaks in the measure series. Structural breaks are the manifestation of a change in the process generating rates of patient harm, and were examined because they are points at which any trend in measure changes that pre-existed PfP would have changed. Thus, to the extent that patient harms exhibit significant structural breaks toward greater reductions, there is evidence that aspects of patient safety changed in meaningful ways, producing results during the PfP campaign.

It is important to note that the baseline periods and length of each measure reported by the HENs varied, both across HENs and across measures within a HEN. In the current analysis, the Evaluation Contractor only assessed the significance of structural breaks during the PfP campaign. This strategy was chosen because the majority of measures reported by the HENs did not include monthly or quarterly data to establish a baseline trend. Rather, the majority of baseline data received from the HENs consisted of single data points consisting of data aggregated over large time periods such as an entire year. Due to this reporting structure, the Evaluation Contractor is not able to determine if any trends in the data are the continuation of pre-existing trends, or due to a structural break coinciding with the start of the HEN contracts.

Finally, having examined the data for the presence of significant structural breaks, the Evaluation Contractor examined the timing and direction of structural breaks to better understand the contexts of quality improvements in patient safety that were realized during the course of the PfP campaign.

Results for All AEAs

Figure 4-1—Change from Baseline to End of Series and ITS Results for Outcome Measures, All AEAs



Source: The Evaluation Contractor's analysis of HEN-level data.

Across all AEAs, the PfP HENs submitted 1,483 outcome measures, as shown in Figure 4-1. Of these outcomes, 56.91 percent of the measures (n = 844) were widely and consistently reported. Widely reported measures have at least 60 percent of aligned hospitals in the HEN reporting for at least 50 percent of the observations in the series. Consistently reported measures have hospital counts that are within, at most, 15 percent of the maximum hospital count in the series.[4-1]

Across all measures, 43.16 percent of measures (n = 640) improved by at least 17.6 percent—or 10 percent for readmissions—representing the reduction in harm required to meet the campaign goal of reducing preventable harm by 40 percent.[4-2] Beyond the 17.6 percent/10 percent goal, 21.85 percent of measures (n = 324) achieved a 40 percent reduction, or 20 percent for readmissions. Thus, nearly half of the measures tested exhibited harm reductions on the scale of the PfP campaign goals.

Of the 1,483 outcome measures examined, 1,450 contained sufficient information to determine whether the measures exhibited a statistically significant change. These measures could be classified as either improving or worsening if the results were significant, or not changing if the results were not significant. Out of the 1,450 measures tested, 42.48 percent (n = 616) exhibited significant improvement while 11.03 percent (n = 160) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 3.85 to 1, indicating that outcomes were nearly four times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 46.48 percent (n = 674) of the measures exhibited no significant changes from the beginning of reporting to the end.

4-1 For measures where the hospital counts fluctuated by more than 15 percent, the Evaluation Contractor estimated a regression equation where the outcome was the measure rate, which was predicted by a linear time trend and a dummy variable equal to 0 if an observation was within 15 percent of the maximum hospital count, and 1 if the observation as not within 15 percent. Where the results indicated that there was no significant relationship between the outcome rate and the dummy variable for hospital counts outside the 15 percent threshold, the measure was defined as consistently reported.

4-2 The Agency for Healthcare Research and Quality (AHRQ) estimates indicate that 44 percent of patient harms are preventable on average. Therefore, 40 percent reduction of 44 percent of preventable harm equated to an absolute reduction of (0.40*0.44) = 17.6 percent.

Among the 616 measures that exhibited significant improvement, the average level of improvement was a 36.13 percent reduction in patient harms. This is more than twice the percentage reduction required to achieve the campaign goal of a 40 percent reduction in the 44 percent of harms that are preventable (i.e., 17.6 percent reduction in the overall rate of patient harms). Therefore, when measures improved in the PfP campaign, the improvements tended to be substantial reductions in patient harms. In contrast, when measures exhibited significant worsening, the average increase in patient harms was 78.52 percent.[4-3] Although the worsening measures exhibited relatively larger rate increases in harms, some of these increases may reflect moderate changes in harms with low base rates at baseline.

Across all AEAs, there were 1,265 measures with at least eight data points, the minimum threshold used in this report for ITS analysis.[4-4] The ITS analysis was applied to test each eligible structural break point from the fifth through the *t*-4th observation (i.e., the fourth from the last data point). This method allowed for at least four observations before and four observations after the tested break point. A structural break is indicated either by a significant change in the trend of the measure from the pre-intervention period to the post-intervention period, or by an instantaneous shift in the level of the rate at the beginning of the post-intervention period.[4-5] Breaks in trend can either be toward more improvement (or less worsening), or toward less improvement (or greater worsening). A structural break in the form of a level shift upward or downward may be accompanied by a change in trend, or may exhibit no change in trend. This type of change in level, without an accompanying change in trend, is also captured in the assessment of overall change in level.

Among the ITS models estimated, 22.37 percent (n = 283) of measures exhibited at least one structural break point associated with a change in trend toward greater improvement. In contrast, 15.97 percent (n = 202) exhibited structural breaks associated with changes in trend toward less improvement. The majority of measures (61.66 percent) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 17,359 tests performed across all measures, with 21.86 percent exhibiting significant structural breaks: 10.27 percent toward greater improvements, and 11.59 toward lesser improvement. The results presented in Figure 4-1, indicate that a larger number of measures changed during the course of the campaign, with approximately 40.08 percent more measures showing trend

4-3 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an obstetrical early elective delivery (OB-EED) rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EEDs from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

4-4 The research literature presents varying estimates regarding the minimum number of data points required to estimate an ITS model, from 16 observations (Penfold and Zhang 2013) to 50 observations (Wagner et al. 2002). Zhang et al. (2011) conducted a power analysis of segmented regression concluding that situations with fewer than 18 data points should be avoided due to insufficient power to detect changes in the series. In the current analysis, only 42.01 percent of the 1,483 outcome measures (n = 623) have at least 18 data points. The proportion of measures meeting this criteria by AEA ranged from a low of 28.57 percent (n = 46 out of 161) in pressure ulcers, to a high of 70.11 percent (n = 61 out of 87 measures). In all AEAs except readmissions, the percentage of measures with at least 18 data points was less than 55 percent. With such a large portion of measures falling below the minimum threshold for ITS analysis, the Evaluation Contractor relaxed the minimum number of observations to 8, expecting that the proportion of measures exhibiting no significant break will be higher than would otherwise be detected with longer data series. However, where the ITS analysis identifies breaks in short time series with reduced power, one can be reasonably confident that significant results are not in error.

4-5 In the ITS framework, a measure must have sufficient observations during both the pre-intervention and post-intervention periods to establish a reliable trend in the series. For this reason, monthly measures could not be tested for significant breaks during Q1 2012 and Q4 2014, which represent the first and last quarters of the campaign. In contrast, quarterly measures could only be tested for significant breaks in the series between Q1 2013 and Q4 2013.

breaks toward greater improvement than those showing trend breaks toward less improvement.[4-6] Thus, it is clear that these changes in measures are not likely to be due to random chance.

Combining the assessment of the overall change in rate (i.e., improving, no change, worsening), with the assessment of structural breaks in trend (i.e., toward more improvement, no change, toward less improvement) yields nine possible profiles of measure behavior. To ease the interpretation of results, however, the Evaluation Contractor combined profiles into three mutually exclusive categories.

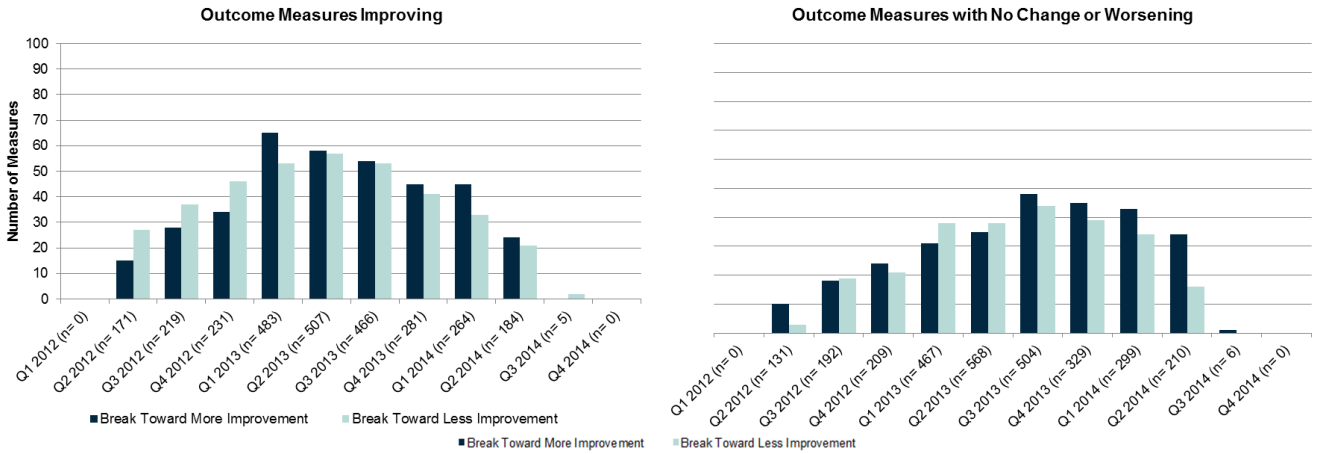
- Measures exhibiting overall improvement in rate and a structural break in trend toward more improvement. This pattern of improvement is most consistent with an intentional improvement in quality.
- Measures exhibiting overall improvement, with either no change in trend or despite a structural break toward less improvement. This pattern of improvement may be consistent with a pre-existing trend, or may be consistent with an improvement reaching a point of diminishing returns, where additional quality improvement is likely to require new and/or additional strategies.
- Measures that exhibited no significant improvement in overall rate or worsening over time, or exhibited structural breaks toward less improvement. This pattern of change provides a less compelling story with respect to quality improvements reducing patient harms. Therefore, the Evaluation Contractor combined the profiles into a single group for purposes of discussion.

Of the 1,265 measures examined, 11.78 percent (n = 149) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent approximately half of the 22.37 percent of measures with structural breaks toward more improvement, indicating that quality improvements (interventions) resulting in significant improvement in trend were more likely to be associated with overall improvements in measure rates. Another 32.96 percent (n = 417) also improved, but either without a structural break in trend, or with a break toward less improvement.

Having identified measures exhibiting structural breaks in trends toward more improvement, it is useful to examine the timing of such breaks to identify when measures were more likely to exhibit improvements in trend, or worsening in trend. Importantly, the vast majority of measures submitted by the HENs did not include trended data for baseline measurements; the measures were most often submitted with an aggregate data point representing the baseline period. Therefore, the Evaluation Contractor can only effectively examine the timing of structural breaks occurring during the PfP campaign.

[4-6] The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((22.37 – 15.97)/15.97)*100 = 40.08 percent.

Figure 4-2—Timing of Structural Breaks in Trend for HEN-Level Outcomes, All AEAs



Source: The Evaluation Contractor's analysis of HEN-level data.

Figure 4-2 shows two graphs that illustrate the timing of structural breaks in HEN outcome measures, across all AEAs. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are four important results to note in Figure 4-2. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

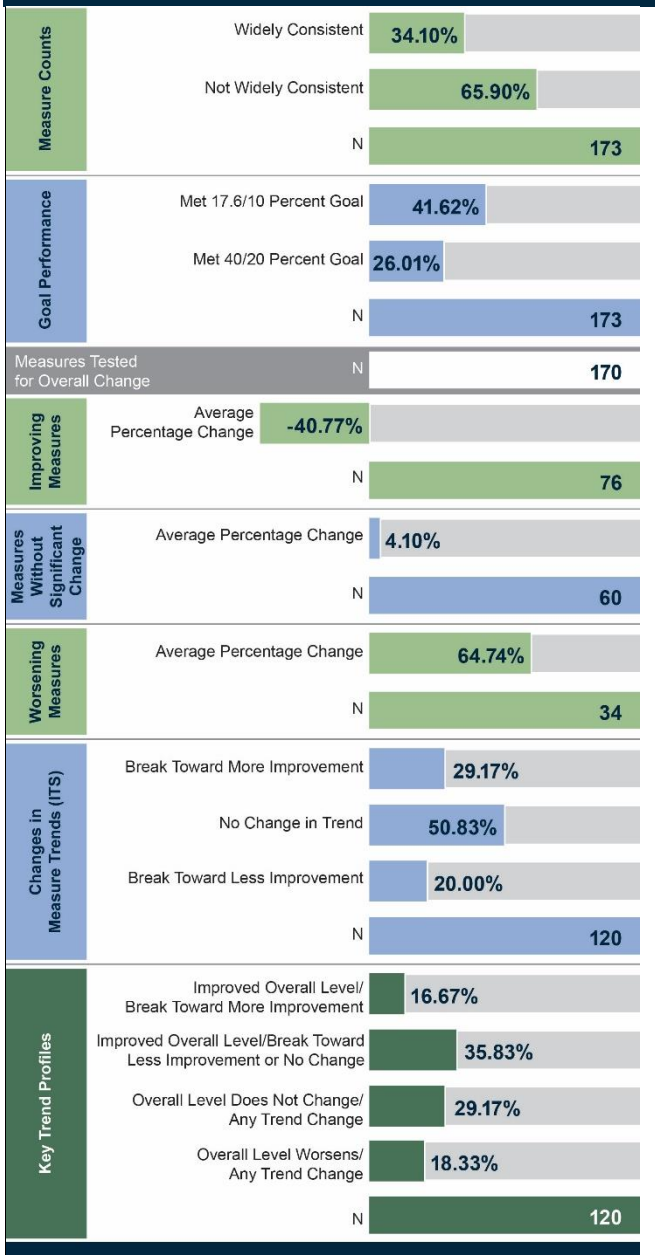
Second, the ratio of breaks toward more improvement and breaks toward less improvement are very consistent overall, with no quarters exhibiting large disparities in the types of breaks identified.

Third, given the consistency in the composition of break type in each quarter, the ITS models identified slightly greater numbers of breaks toward less improvement during the first half of the PfP campaign, and slightly greater numbers of breaks toward more improvement during the latter half of the PfP campaign. This pattern is consistent with a story in which organizational learning develops over time. While quality improvements were made throughout the campaign, the measures are more likely to exhibit breaks toward more improvement later in the campaign, after the HENs and hospitals are likely to have learned how to develop more effective interventions.

Fourth, it is important to note that measures exhibiting overall improvement appear more likely to have begun exhibiting structural breaks toward more improvement sooner than measures that did not exhibit overall improvement. The change in dominant break types for improved measures occurred in Q1 2013, whereas for non-improved measures the change occurred 6 months later, in Q3 2013. This pattern suggests that even among measures that did not exhibit overall improvement during the three years of the PfP campaign, measures were more likely to begin breaking toward improvement during the last 15 months of the campaign. If this pattern was to hold, and the campaign had continued beyond December 2014, one would expect to see more measures exhibit significant overall improvements.

Results for Adverse Drug Events (ADEs)

Figure 4-3—Change from Baseline to End of Series and ITS Results for Outcome Measures, ADE



Source: The Evaluation Contractor's analysis of HEN-level data.

Among ADE outcomes, the PfP HENs submitted 173 outcome measures, as shown in Figure 4-3. Of these outcomes, 34.10 percent of the measures (n = 59) were widely and consistently reported. This topic had the lowest percentage of widely-reported and consistent outcome measures across all AEAs in the PfP campaign, or to put it another way, the highest degree of variation in measures attempted. Therefore, the results presented in this section are likely to be the least representative of the overall HEN performance among all of the topic areas.

Across ADE measures, 41.62 percent of measures (n = 72) improved by at least 17.6 percent. Beyond the 17.6 percent goal, 26.01 percent of measures (n = 45) achieved a 40 percent reduction. Thus, approximately 40 percent of the ADE measures tested exhibited harm reductions on the scale of the PfP campaign goals.

From the 173 outcome measures examined, 170 contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 170 ADE measures tested, 44.71 percent (n = 76) exhibited significant improvement while 20.00 percent (n = 34) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 2.24 to 1, indicating that ADE outcomes were just over two times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 35.29 percent (n = 60) of the measures exhibited no significant changes from the beginning of reporting to the end.

Among the 76 ADE measures that exhibited significant improvement, the average level of improvement was a 40.77 percent reduction in patient harms. This is more than twice the percentage reduction required to achieve a 40 percent reduction in patient harms, assuming 44 percent are preventable. Therefore, when ADE measures improved in the PfP campaign, the improvements tended to be substantial reductions in patient harms. In contrast, when measures exhibited significant worsening, the average increase in patient harms was 64.74 percent.[4-7]

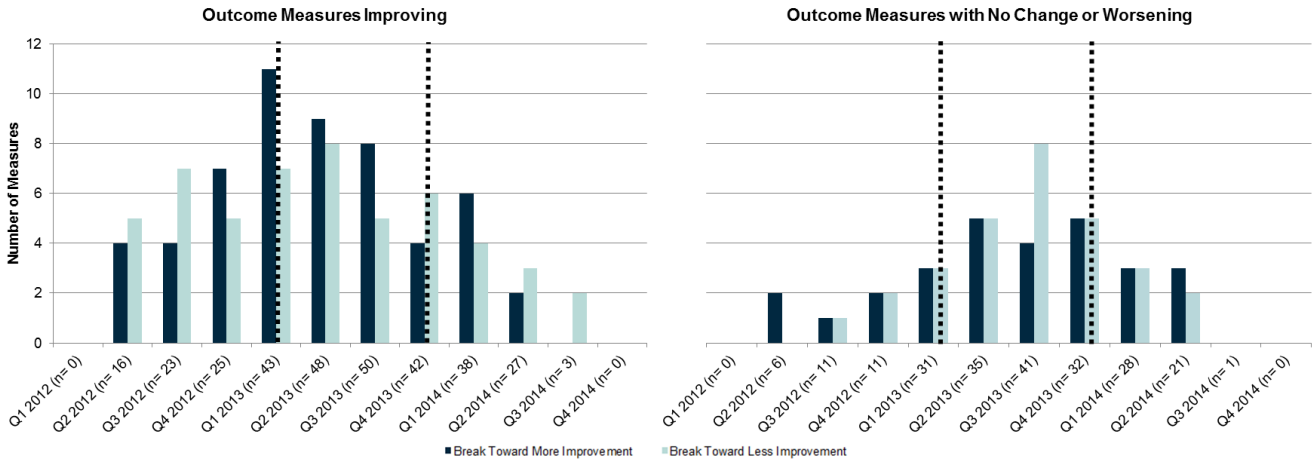
Thus, measures of patient harms that worsened significantly exhibited rate increases of substantial amounts, although some of these increases may reflect moderate changes in harms with low base rates at baseline.

4-7 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EEDs from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

Across ADE outcomes, there were 120 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for ADE measures, 29.17 percent (n = 35) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 20.00 percent (n = 24) exhibited structural breaks associated with changes in trend toward less improvement. The majority of measures (50.83 percent) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 1,542 tests performed across all measures, with 27.43 percent exhibiting significant structural breaks: 12.19 percent toward greater improvements, and 15.24 toward lesser improvement. The results presented in Figure 4-3, indicate a larger number of measures changed during the course of the campaign, with approximately 45.85 percent more measures showing trend breaks toward greater improvement than those showing trend breaks toward less improvement.[4-8] Thus, these changes in measure rates are not likely to be due to random chance, and indicate that processes producing ADEs were changing toward improvement.

Of the 120 measures examined with ITS, 16.67 percent (n = 20) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent over half of the 29.17 percent of measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were more likely to be associated with overall improvements in measure rates. Another 35.83 percent (n = 43) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-4—Timing of Structural Breaks in Trend for HEN-Level ADE Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.
Note: The dashed lines in the figure represent the following campaign-level events: push for data and workshop series for ADEs (January 2013–March 2013) and request that all HENs begin reporting on three ADE areas: glycemic control, opioids, and anticoagulants (December 2013).

Figure 4-4 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for ADE. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-4. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points

4-8 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((29.17 – 20.00)/20.00)*100 = 45.85 percent.

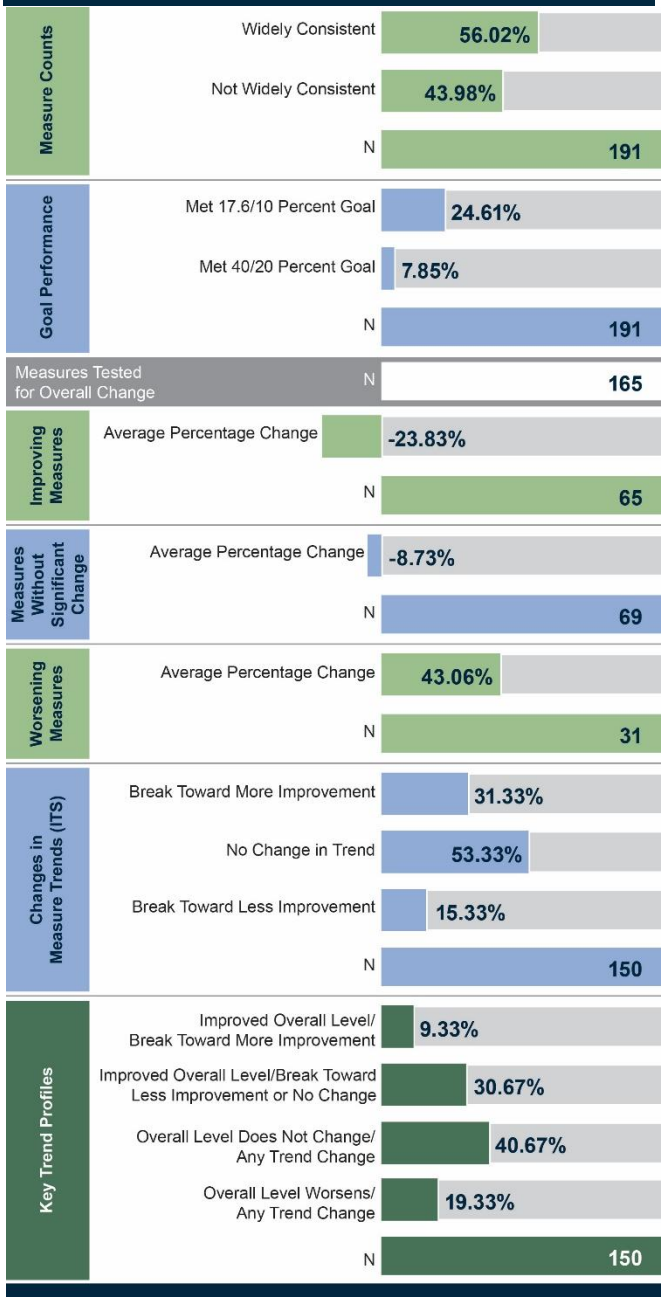
during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are not consistent between measures that improved and those that did not. For measures that improved, more breaks toward more improvement occurred between Q4 2012 and Q3 2013, as well as in Q1 2014. This pattern aligns with the PfP campaign pushes for ADE data and workshop series in Q1 2013, and the refocusing of ADE on three high-harm areas in Q4 2013: anticoagulants, glycemic control, and opioids. In contrast, among measures that did not improve overall, there were more breaks toward more improvement than toward less improvement only in Q2 2012 and Q2 2014, although in several quarters the breaks were evenly distributed between breaks toward more and less improvement. The timing of these changes does not align well with campaign-level events pushed in ADE.

Third, it is important to note that measures exhibiting overall improvements appear more likely to have begun exhibiting structural breaks toward more improvement sooner than measures that did not exhibit overall improvement. Among measures that improved overall, the peak of structural breaks toward more improvement occurred in Q1 2013, whereas the peak is between Q2 and Q4 2013 for measures that did not improve overall. In contrast to the pattern among all AEA outcomes that did not improve overall, there is no pattern of structural breaks for ADE measures that did not improve that indicates an increased likelihood of moving toward improvements during the PfP campaign. If this pattern was to hold, and the campaign had continued beyond December 2014, one would not necessarily expect to see more ADE measures exhibit significant overall improvements.

Results for Catheter-Associated Urinary Tract Infections (CAUTI)

Figure 4-5—Change from Baseline to End of Series and ITS Results for Outcome Measures, CAUTI



Source: The Evaluation Contractor's analysis of HEN-level data.

Among CAUTI outcomes, the PfP HENs submitted 191 outcome measures, as shown in Figure 4-5. Of these outcomes, 56.02 percent of the measures (n = 107) were widely and consistently reported.

Across CAUTI measures, 24.61 percent of measures (n = 47) improved by at least 17.6 percent. This is the smallest proportion of measures achieving the 17.6 percent goal across all AEAs. Beyond the 17.6 percent goal, 7.85 percent of measures (n = 15) achieved a 40 percent reduction. This is less than half of the percentage of measures achieving the 40 percent goal for the next closest AEA. Thus, only one quarter of the CAUTI measures tested exhibited harm reductions on the scale of the PfP campaign goals.

From the 191 outcome measures examined, 165 contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 165 CAUTI measures tested, 39.39 percent (n = 65) exhibited significant improvement while 18.79 percent (n = 31) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 2.08 to 1, indicating that CAUTI outcomes were just over two times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 41.82 percent (n = 69) of the measures exhibited no significant changes from the beginning of reporting to the end.

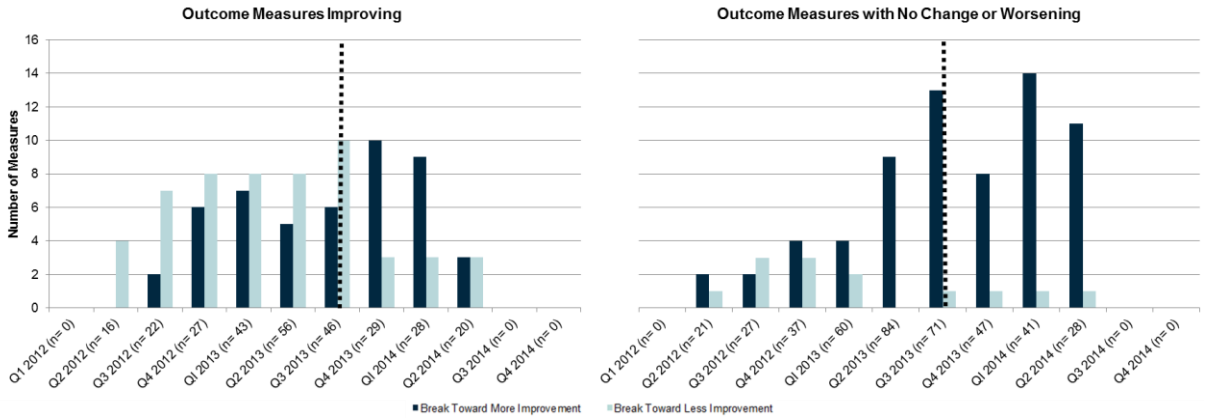
Among the 65 CAUTI measures that exhibited significant improvement, the average level of improvement was a 23.83 percent reduction in patient harms. This is 6 percentage points greater than the reduction required to achieve a 40 percent reduction in patient harms, assuming 44 percent are preventable. Therefore, among the CAUTI measures that improved in the PfP campaign, the improvements were not trivial in nature. The majority of improvements among CAUTI outcome measures are associated with reductions in device utilization ratios, which improved among the majority of HENs reporting. In contrast, when measures exhibited significant

worsening, the average increase in patient harms was 43.06 percent.[4-9] Thus, measures of patient harms that worsened significantly exhibited rate increases of substantial amounts, although some of these increases may reflect moderate changes in harms with low base rates at baseline.

Across CAUTI outcomes, there were 150 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for CAUTI outcomes, 31.33 percent (n = 47) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 15.33 percent (n = 23) exhibited structural breaks associated with changes in trend toward less improvement. The majority of measures (53.33 percent) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 1,906 tests performed across all measures, with 22.82 percent exhibiting significant structural breaks: 13.27 percent toward greater improvements, and 9.55 toward lesser improvement. The results presented in Figure 4-5, indicate a larger number of measures changed during the course of the campaign, with approximately 104.37 percent more measures showing trend breaks toward greater improvement than those showing trend breaks toward less improvement.[4-10] Thus, these changes in measures are not likely to be due to random chance, and indicate that processes producing CAUTIs were changing toward improvement, even though the CAUTI rates themselves did not exhibit dramatic improvements.

Of the 150 measures examined with ITS, 9.33 percent (n = 14) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent one-third of the 31.33 percent of measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were not more likely to be associated with overall improvements in measure rates. Another 30.67 percent (n = 46) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-6—Timing of Structural Breaks in Trend for HEN-Level CAUTI Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.
Note: The dashed line in the figure represents the following campaign-level event: Stop CAUTI event for HENs and hospitals (July 2013 and September 2013).

4-9 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EED from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

4-10 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((31.11– 15.33)/15.33)*100 = 104.37 percent.

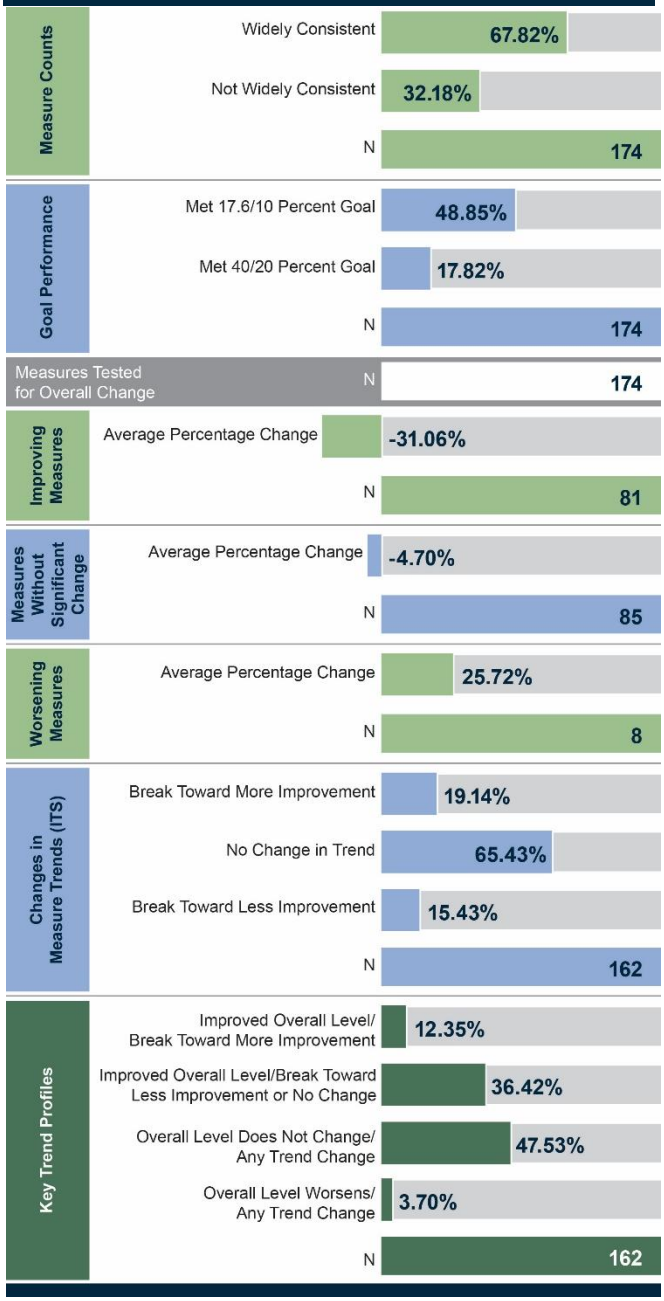
Figure 4-6 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for CAUTI. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-6. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are not consistent between measures that improved and those that did not. For measures that improved, more breaks toward less improvement occurred between Q2 2012 and Q3 2013. During Q4 2013 and Q1 2014; however, this pattern reversed, and a larger number of breaks toward more improvement were identified compared to breaks toward less improvement. This pattern aligns with the PfP campaign *Stop CAUTI* events in July 2013 and September 2013. In contrast, among measures that did not improve overall, there were more breaks toward more improvement than toward less improvement in every quarter except for Q3 2012. Beginning in Q2 2013, however, the number of breaks toward more improvement identified increased dramatically, even though the number of measures being tested did not increase by the same amount. This result suggests that CAUTI rates among measures that did not improve were nonetheless changing in a direction suggesting incremental gains, or at least a reduction in the worsening of these outcome rates. The timing of these changes does not align particularly well with the campaign-level *Stop CAUTI* event.

Third, it is important to note that measures exhibiting overall improvement appear to have begun exhibiting structural breaks toward more improvement at approximately the same time as measures that did not exhibit overall improvement. For both groups of measures, the peak of structural breaks toward more improvement occurred in Q4 2013 and Q1 2014. In contrast, among measures that improved overall, there is a substantially larger number of breaks toward less improvement between Q2 2012 and Q3 2013, suggesting that several measures may have reached a floor beyond which additional improvements were attenuated. In contrast, among the measures that did not improve, fewer structural breaks toward less improvement were identified, suggesting that CAUTI rates were more likely to remain stable or continue worsening until the middle of the campaign, when the number of breaks toward more improvement increased. This pattern is consistent with a topic area in which the HENs exhibited difficulty making major gains in reducing patient harms, although trends in CAUTI outcomes were more likely to be changing in ways consistent with greater improvements toward the end of the PfP campaign. This implies that continued work in this area may be likely to exhibit a greater prevalence of substantive improvements in the future, assuming that efforts expended by the HENs in 2014 continue.

Results for Central Line-Associated Blood Stream Infections (CLABSI)

Figure 4-7—Change from Baseline to End of Series and ITS Results for Outcome Measures, CLABSI



Source: The Evaluation Contractor's analysis of HEN-level data.

Among CLABSI outcomes, the PfP HENs submitted 174 outcome measures, as shown in Figure 4-7. Of these outcomes, 67.82 percent of the measures (n = 118) were widely and consistently reported. This is about 10 percentage points higher than for all outcomes across AEAs.

Across CLABSI measures, 48.85 percent of measures (n = 85) improved by at least 17.6 percent. This is the fourth largest proportion of measures achieving the 17.6 percent goal across all AEAs. Beyond the 17.6 percent goal, 17.82 percent of measures (n = 31) achieved a 40 percent reduction. This is the fifth highest proportion of measures achieving this goal across all AEAs. Thus, just under half of the CLABSI measures tested exhibited harm reductions on the scale of the PfP campaign goals.

All 174 outcome measures examined contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 174 CLABSI measures tested, 46.55 percent (n = 81) exhibited significant improvement while 4.60 percent (n = 8) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 10.13 to 1, indicating that CLABSI outcomes were more than 10 times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 48.85 percent (n = 85) of the measures exhibited no significant changes from the beginning of reporting to the end.

Among the 81 CLABSI measures that exhibited significant improvement, the average level of improvement was a 31.06 percent reduction in patient harms. This is nearly twice the reduction required to achieve a 40 percent reduction in patient harms, assuming 44 percent are preventable. Therefore, among the CLABSI measures that improved in the PfP campaign, the improvements were substantial. The majority of improvements among CLABSI outcome measures are associated with reductions in both device utilization ratios and infection rates. In contrast, when measures exhibited significant worsening, the average increase in patient harms was only 25.72 percent.[4-11] Thus, measures of

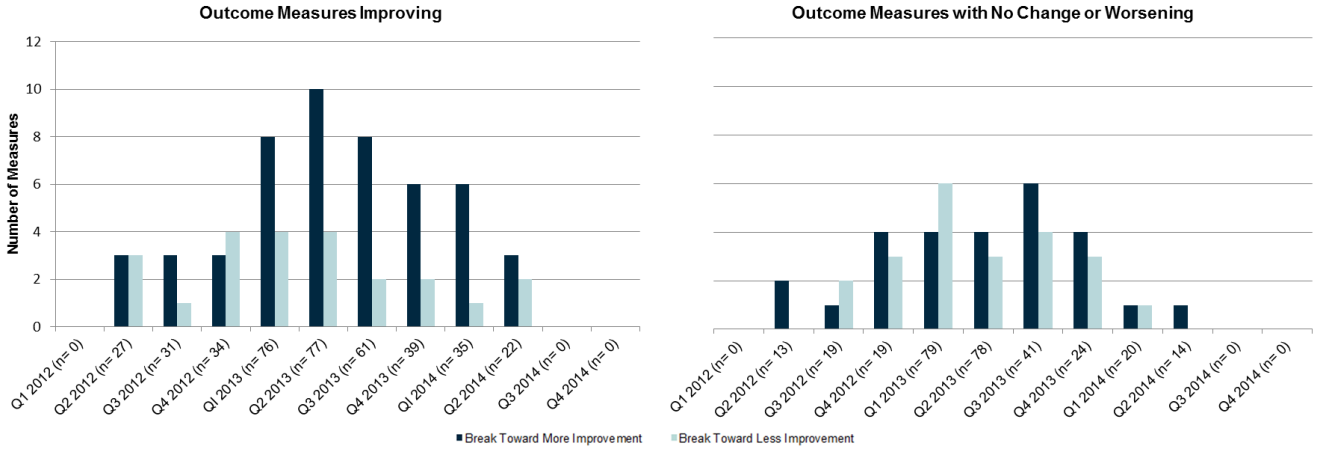
4-11 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EEDs from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

patient harms that worsened significantly exhibited rate increases that were more than trivial, yet not as substantial as changes in other AEAs, although some of these increases may reflect moderate changes in harms with low base rates at baseline.

Across CLABSI outcomes, there were 162 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for CLABSI outcomes, 19.14 percent (n = 31) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 15.43 percent (n = 25) exhibited structural breaks associated with changes in trend toward less improvement. The majority of measures (65.43 percent) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 1,801 tests performed across all measures, with 16.10 percent exhibiting significant structural breaks: 10.77 percent toward greater improvements, and 5.33 toward lesser improvement. The results presented in Figure 4-7, indicate a larger number of measures changed during the course of the campaign, with approximately 24.04 percent more measures showing trend breaks toward greater improvement than those showing trend breaks toward less improvement.[4-12] Thus, these changes in measures are not likely to be due to random chance, and indicate that processes producing CLABSI rates, as well as infection rates were changing toward improvement.

Of the 162 measures examined with ITS, 12.35 percent (n = 20) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent two-thirds of the 31 measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were more likely to be associated with overall improvements in measure rates. Another 36.42 percent (n = 59) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-8—Timing of Structural Breaks in Trend for HEN-Level CLABSI Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.

4-12 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((19.14– 15.43)/15.43)*100 = 24.04 percent.

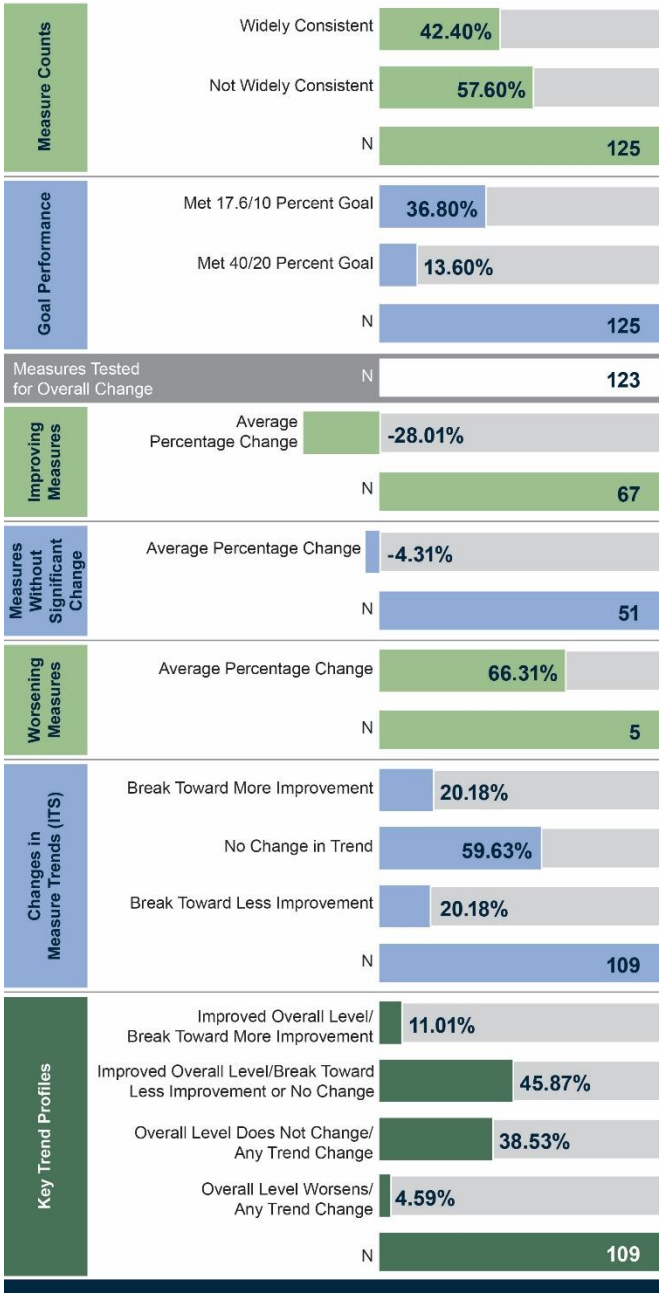
Figure 4-8 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for CLABSI. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-8. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are not consistent between measures that improved and those that did not. For measures that improved, more breaks toward more improvement occurred between Q1 2013 and Q2 2014. Among measures that did not improve overall, the ratio of structural breaks toward more or less improvement only slightly favored breaks toward more improvement during the PfP campaign. This pattern is consistent with a period during 2012 wherein CLABSI measures were approximately equally likely to exhibit breaks toward more improvement as toward less improvement. Greater progress was made during 2013 and the first half of 2014.

Third, it is important to note that measures exhibiting overall improvement appear to have begun exhibiting structural breaks toward more improvement earlier than measures that did not exhibit overall improvement. For improved measures, the peak of structural breaks toward more improvement occurred in Q2 2013. In contrast, among measures that did not improve overall, the number of structural breaks toward improvement is roughly constant through 2013, with the peak occurring in Q3 2013. This pattern is consistent with a topic area in which there was initial difficulty in making improvements in CLABSI overall, but once progress began, the proportion of measures exhibiting improvements relative to worsening trends increased substantially.

Results for Falls

Figure 4-9—Change from Baseline to End of Series and ITS Results for Outcome Measures, Falls



Source: The Evaluation Contractor's analysis of HEN-level data.

Among falls outcomes, the PfP HENs submitted 125 outcome measures, as shown in Figure 4-9. Of these outcomes, 42.40 percent of the measures (n = 53) were widely and consistently reported. This is less than the percentage of widely-reported and consistent measures for all outcomes across AEAs. Thus the outcomes for falls are not as representative of HEN performance as outcomes in some other AEAs

Across falls measures, 36.80 percent of measures (n = 46) improved by at least 17.6 percent. Beyond the 17.6 percent goal, 13.60 percent of measures (n = 17) achieved a 40 percent reduction. Thus, just over one-third of the falls measures tested exhibited harm reductions on the scale of the PfP campaign goals.

Of the 125 outcome measures examined, 123 contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 123 falls measures tested, 54.47 percent (n = 67) exhibited significant improvement while 4.07 percent (n = 5) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 13.40 to 1, indicating that falls outcomes were more than 13 times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 41.46 percent (n = 51) of the measures exhibited no significant changes from the beginning of reporting to the end.

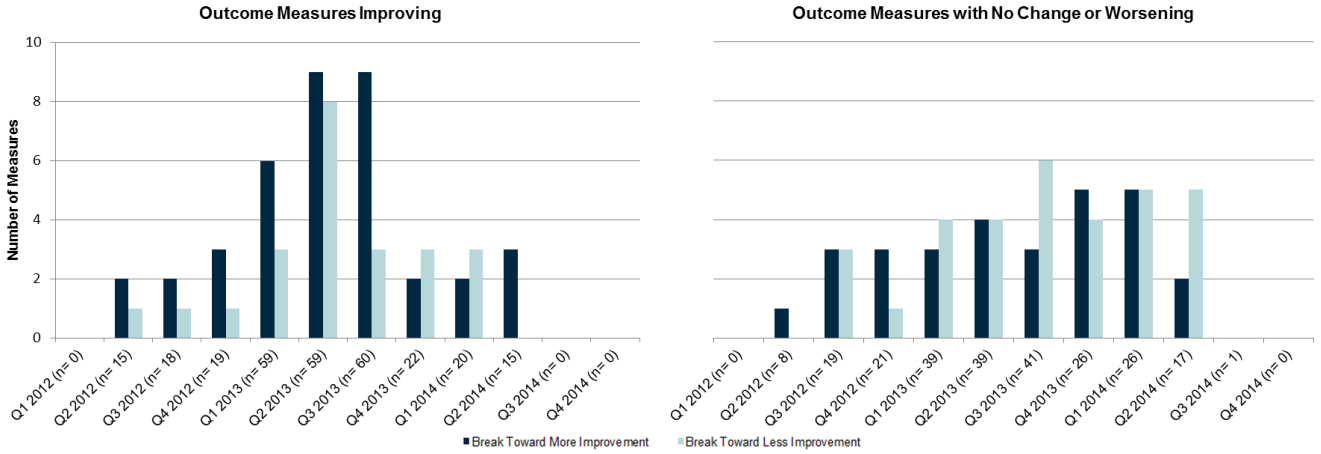
Among the 67 falls measures that exhibited significant improvement, the average level of improvement was a 28.01 percent reduction in patient harms. This is 10 percentage points larger than the reduction required to achieve a 40 percent reduction in patient harms, assuming 44 percent are preventable. Therefore, among the falls measures that improved in the PfP campaign, the improvements were substantial. In contrast, when measures exhibited significant worsening, the average increase in patient harms was 66.31 percent.[4-13] Thus, measures of patient harms that worsened significantly exhibited rate increases that were substantial, although some of these increases may reflect moderate changes in harms with low base rates at baseline.

4-13 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EED from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

Across falls outcomes, there were 109 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for falls outcomes, 20.18 percent (n = 22) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 20.18 percent (n = 22) exhibited structural breaks associated with changes in trend toward less improvement. The majority of measures (59.63 percent) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 1,513 tests performed across all measures, with 24.19 percent exhibiting significant structural breaks: 10.51 percent toward greater improvements, and 13.68 toward lesser improvement. The results presented in Figure 4-9, indicate just under half of the measures changed during the course of the campaign, with approximately equal proportions of measures showing trend breaks toward greater improvement as those showing trend breaks toward less improvement.[4-14] Thus, these changes in measures are relatively balanced, and may indicate that processes producing falls rates were changing toward improvements in some HENs, and not in others.

Of the 109 measures examined with ITS, 11.01 percent (n = 12) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent just over half of the 22 measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were only mildly more likely to be associated with overall improvements in measure rates. Another 45.87 percent (n = 50) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-10—Timing of Structural Breaks in Trend for HEN-Level Falls Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.

Figure 4-10 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for falls. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-10. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points

4-14 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((20.18– 20.18)/20.18)*100 = 0 percent difference.

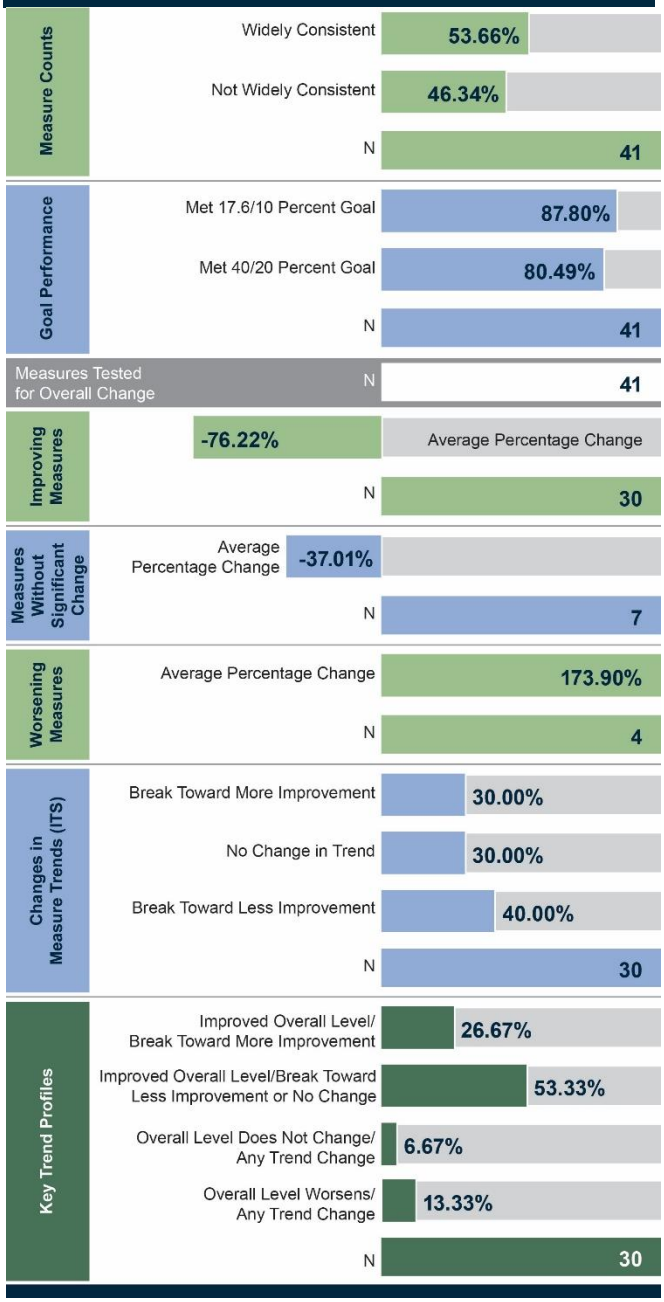
during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are not consistent between measures that improved and those that did not. For measures that improved, more breaks toward more improvement occurred between Q2 2012 and Q3 2013. Among measures that did not improve overall, the ratio of structural breaks toward more or less improvement only slightly favored breaks toward less improvement during the PfP campaign. This pattern is consistent with a period during 2012 and 2013 wherein falls measures were more likely to exhibit improving trends, with fewer measures exhibiting significant breaks toward either more or less improvement during the last year of the PfP campaign.

Third, it is important to note that measures exhibiting overall improvement appear to have begun exhibiting structural breaks toward more improvement earlier than measures that did not exhibit overall improvement. For improved measures, the peak of structural breaks toward more improvement occurred in Q2 and Q3 2013. In contrast, among measures that did not improve overall, the number of structural breaks toward improvement increases only slightly through 2013, with a small peak occurring in Q4 2013 and Q1 2014. This pattern is consistent with a topic area in which there was initial success in making improvements in falls overall. Once initial gains were made and measures were trending toward more improvement in the first half of 2013, fewer measures exhibited structural breaks in either direction.

Results for Obstetrical Early Elective Deliveries (OB-EED)

Figure 4-11—Change from Baseline to End of Series and ITS Results for Outcome Measures, OB-EED



Source: The Evaluation Contractor's analysis of HEN-level data.

Among OB-EED outcomes, the PfP HENs submitted 41 outcome measures, as shown in Figure 4-11. Of these outcomes, 53.66 percent of the measures (n = 22) were widely and consistently reported.

Across OB-EED measures, 87.80 percent of measures (n = 36) improved by at least 17.6 percent. Beyond the 17.6 percent goal, 80.49 percent of measures (n = 33) achieved a 40 percent reduction. Thus, nearly 9 out of 10 OB-EED measures tested exhibited harm reductions on the scale of the PfP campaign goals.

All 41 OB-EED measures contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 41 OB-EED measures tested, 73.17 percent (n = 30) exhibited significant improvement while 9.76 percent (n = 4) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 7.50 to 1, indicating that OB-EED outcomes were more than seven times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 17.07 percent (n = 7) of the measures exhibited no significant changes from the beginning of reporting to the end.

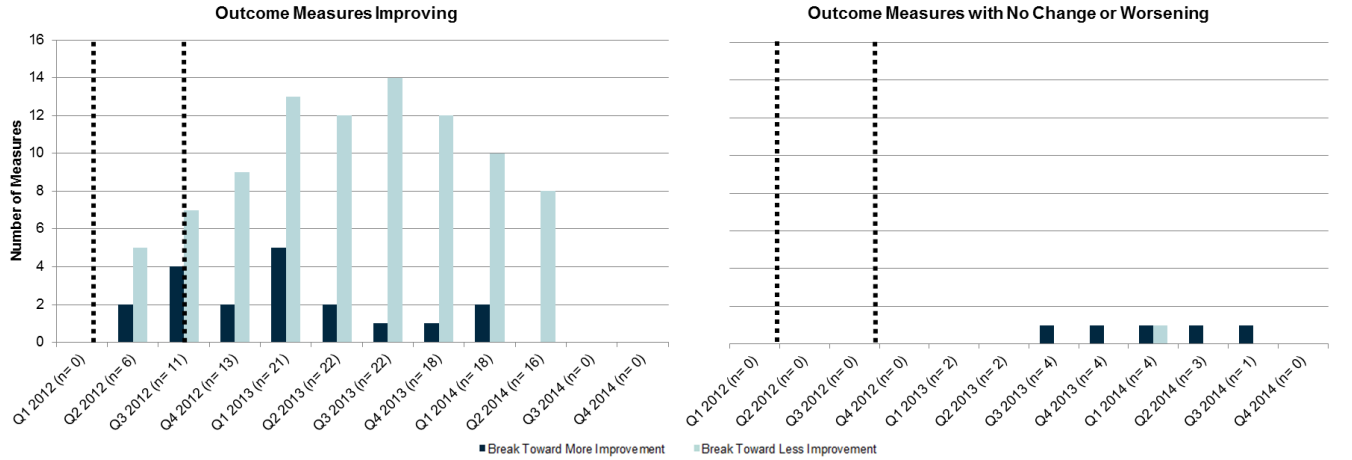
Among the 30 OB-EED measures that exhibited significant improvement, the average level of improvement was a 76.22 percent reduction in OB-EEDs, substantially larger than the required improvement to meet the PfP campaign goals. In contrast, when measures exhibited significant worsening, the average increase in OB-EED was 173.90 percent.[4-15] Thus, measures of patient harms that worsened significantly exhibited rate increases that were substantial, although some of these increases may reflect moderate changes in harms with low base rates at baseline.

4-15 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EED from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

Across OB-EED outcomes, there were 30 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for OB-EED outcomes, 30 percent (n = 9) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 40 percent (n = 12) exhibited structural breaks associated with changes in trend toward less improvement. Only one-third of the measures (n = 9) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 481 tests performed across all measures, with 62.16 percent exhibiting significant structural breaks: 13.31 percent toward greater improvements, and 48.86 toward lesser improvement. The results presented in Figure 4-11, indicate that nearly three-quarters of the measures changed during the course of the campaign, with slightly more measures showing trend breaks toward less improvement as those showing trend breaks toward more improvement.[4-16] This can be explained by the predominant method used to reduce the rate of OB-EEDs: hard-stop policies. Large reductions in OB-EEDs were observed among HENs implementing hard-stop policies, with measure trends flattening out as the rates approach zero, and resulting in a structural break toward less improvement. Thus, these changes in measures are not unexpected and indicate that processes producing OB-EED rates were changing toward improvements in many HENs.

Of the 30 measures examined with ITS, 26.67 percent (n = 8) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent nearly all of the 9 measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were more likely to be associated with overall improvements in measure rates. Another 53.33 percent (n = 16) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-12—Timing of Structural Breaks in Trend for HEN-Submitted OB-EED Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.
Note: The dashed lines in the figure represent the following campaign-level events: Strong Start Initiative announced (February 2012); in-person meeting with HENs with OB-EED commitments for results (August 2012).

Figure 4-12 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for OB-EED. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-12. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete

4-16 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((20.18– 20.18)/20.18)*100 = 0 percent difference.

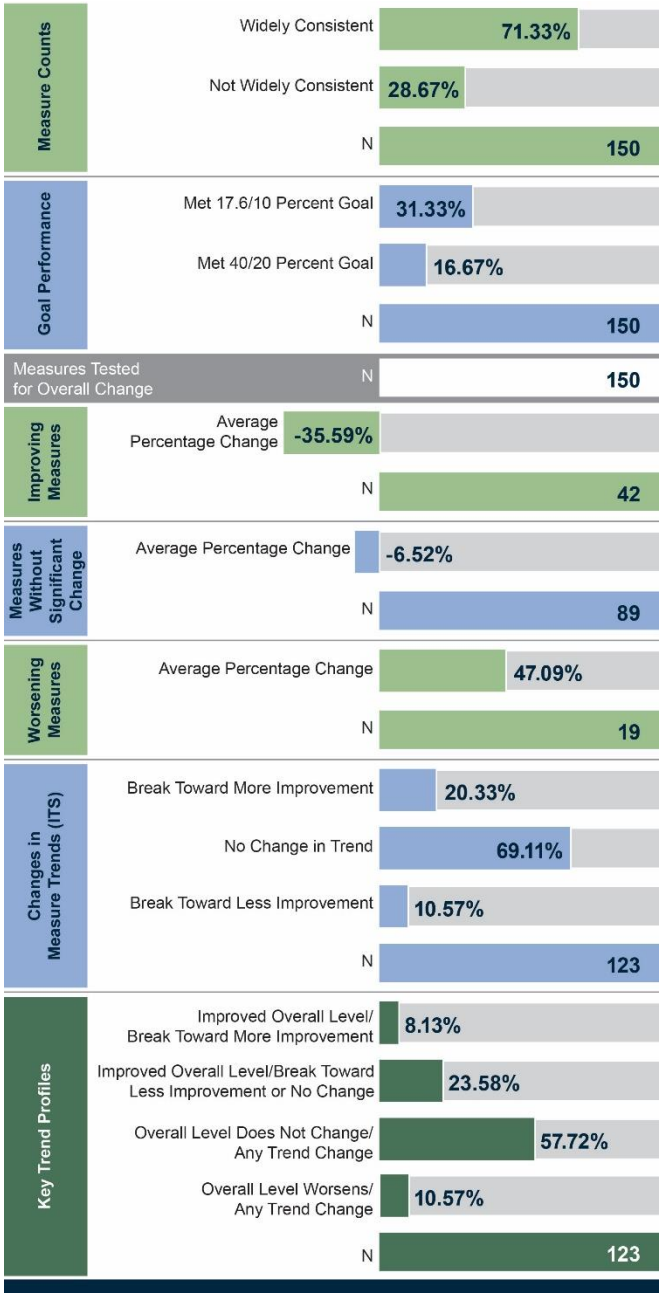
quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are not consistent between measures that improved and those that did not. For measures that improved, more breaks toward less improvement occurred throughout the PfP campaign. This pattern is consistent with the dramatic decline in OB-EED rates associated with hard stop policies implemented largely in 2012, and the subsequent floor effect causing measure trends to flatten out and break toward less improvement. Among measures that did not improve overall, the ratio of structural breaks toward more or less improvement only slightly favored breaks toward more improvement; however, these breaks only occurred for limited measures and only between Q3 2013 and Q3 2014. This pattern is consistent with late improvements being made in a limited number of measures, which are likely to have already low rates.

Third, it is important to note that measures exhibiting overall improvement appear to have begun exhibiting structural breaks toward less improvement earlier than measures that did not exhibit overall improvement. For improved measures, the structural breaks toward less improvement began occurring in Q2 2012, and continued for the next 2 years. In contrast, among measures that did not improve overall, the number of structural breaks toward improvement increases only in Q3 2013, with a single measure exhibiting such a break for the next four quarters. This pattern is consistent with a topic area in which there was large success in making improvements in OB-EED from nearly the beginning of the campaign. This coincides with the implementation of the Strong Start Initiative by the Centers for Medicare & Medicaid Services (CMS) in February 2012, and the HEN commitments to reduce OB-EEDs in August 2012. At the same time that many HENs were implementing hard-stop policies, there were large and abrupt declines in OB-EED outcomes, resulting in structural breaks toward less improvement when the outcomes neared a rate of zero.

Results for Other Obstetrical Adverse Events (OB-Other)

Figure 4-13—Change from Baseline to End of Series and ITS Results for Outcome Measures, OB-Other



Source: The Evaluation Contractor's analysis of HEN-level data.

Among OB-Other outcomes, the PfP HENs submitted 150 outcome measures, as shown in Figure 4-13. Of these outcomes, 71.33 percent of the measures (n = 107) were widely and consistently reported. This is the second highest percentage of widely-reported and consistent measures across all AEAs, and indicates that the results in this section are more likely to be representative of the HEN performance overall than for other AEAs.

Across OB-Other measures, 31.33 percent of measures (n = 47) improved by at least 17.6 percent. Beyond the 17.6 percent goal, 16.67 percent of measures (n = 25) achieved a 40 percent reduction. Thus, about one-third of OB-Other measures tested exhibited harm reductions on the scale of the PfP campaign goals.

All 150 OB-Other measures contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 150 OB-Other measures tested, 28.00 percent (n = 42) exhibited significant improvement, while 12.67 percent (n = 19) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 2.21 to 1, indicating that OB-Other outcomes were two times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 59.33 percent (n = 89) of the measures exhibited no significant changes from the beginning of reporting to the end.

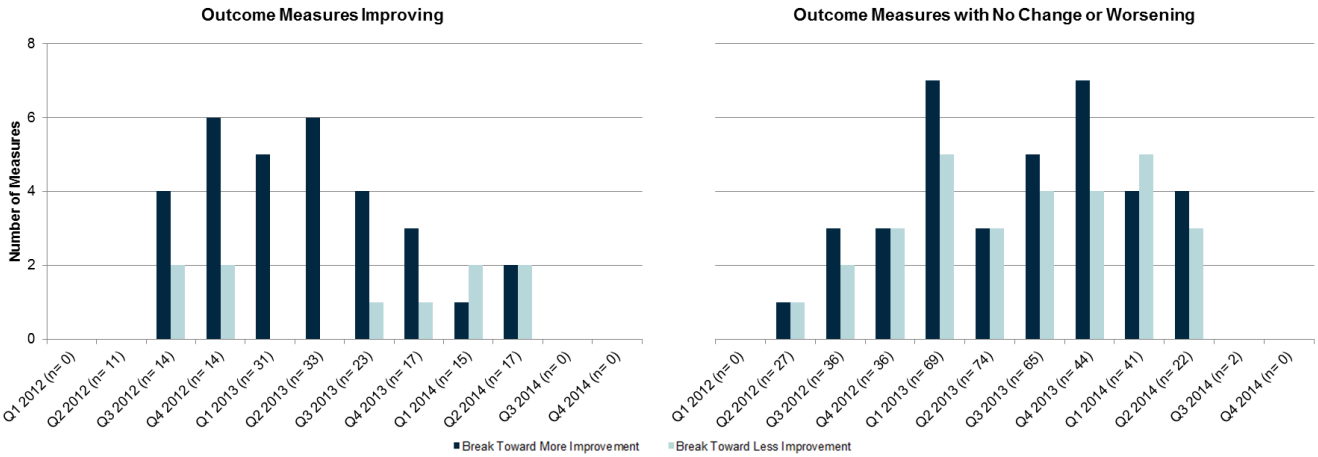
Among the 42 OB-Other measures that exhibited significant improvement, the average level of improvement was a 35.59 percent reduction in patient harms, about twice the required reduction to meet the PfP campaign goals. In contrast, when measures exhibited significant worsening, the average increase in patient harms was 47.09 percent.[4-17] Thus, measures of patient harms that worsened significantly exhibited relatively larger rate increases, although some of these increases may reflect moderate changes in harms with low base rates at baseline.

4-17 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EED from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

Across OB-Other outcomes, there were 123 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for OB-Other outcomes, 20.33 percent (n = 25) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 10.57 percent (n = 13) exhibited structural breaks associated with changes in trend toward less improvement. More than two-thirds of the measures (n = 85) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 1,548 tests performed across all measures, with 15.76 percent exhibiting significant structural breaks: 9.11 percent toward greater improvements, and 6.65 toward lesser improvement. The results presented in Figure 4-13, indicate that nearly one-third of the measures changed during the course of the campaign, with almost twice as many measures showing trend breaks toward more improvement as those showing trend breaks toward less improvement.[4-18] Thus, these changes in measures are not likely due to chance, and indicate that processes producing OB-Other rates were changing toward improvements in the HENs.

Of the 123 measures examined with ITS, 8.13 percent (n = 10) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent 40 percent of the 25 measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were slightly less likely than not to be associated with overall improvements in measure rates. Another 23.58 percent (n = 29) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-14—Timing of Structural Breaks in Trend for HEN-Level OB-Other Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.

Figure 4-14 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for OB-Other. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-14. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

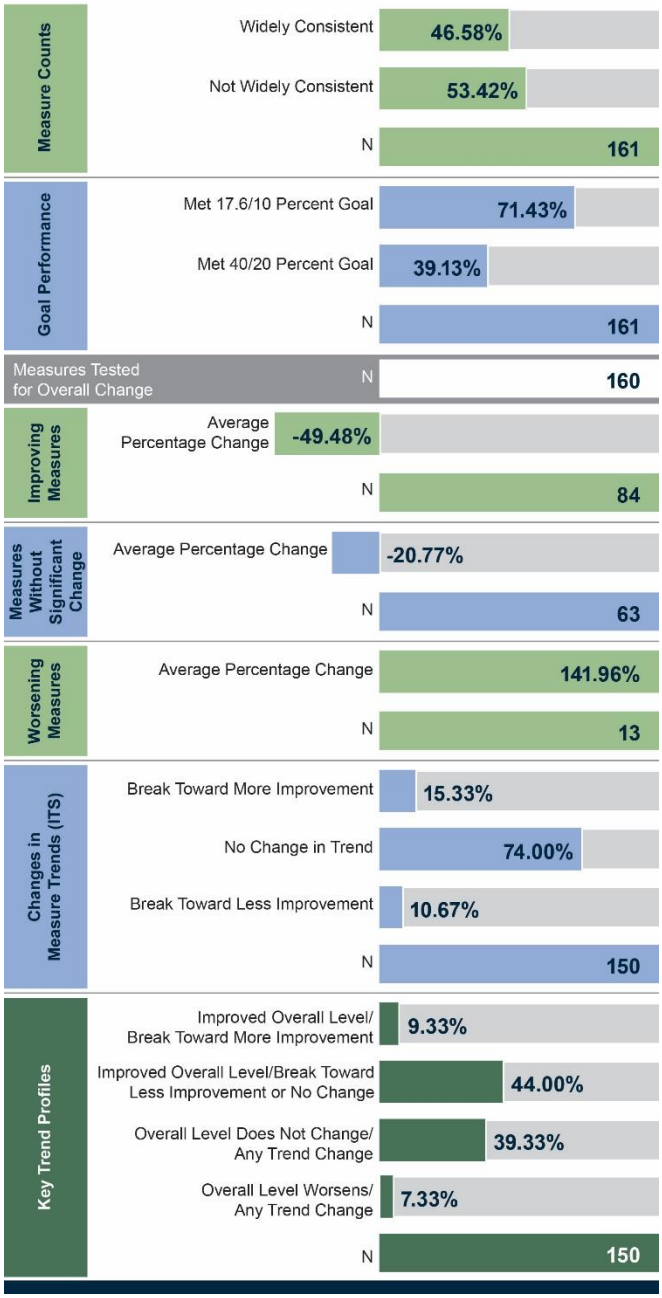
4-18 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((20.33– 10.57)/10.57)*100 = 92.34 percent.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are not consistent between measures that improved and those that did not. For measures that improved, more breaks toward more improvement occurred throughout the PfP campaign. This pattern indicates that among measures that improved during the campaign, there are a greater number of trends that exhibit structural breaks consistent with quality improvement efforts, than in opposition to those efforts. Among measures that did not improve overall, the ratio of structural breaks toward more or less improvement only slightly favored breaks toward more improvement; however, breaks toward more improvement were more likely to occur predominantly during Q1 and Q4 2013.

Third, it is important to note that measures exhibiting overall improvement appear to have begun exhibiting structural breaks toward more improvement earlier than measures that did not exhibit overall improvement. For improved measures, the structural breaks toward more improvement began occurring in Q3 2012, and continued at an elevated level until Q3 2013. In contrast, among measures that did not improve overall, the number of structural breaks toward more improvement increased noticeably in Q1 2013 and Q4 2013, whereas in other quarters there was little difference between types of breaks. Thus, OB-Other measures that exhibited significant improvement were more likely to have also exhibited structural breaks in earlier periods of the PfP campaign.

Results for Pressure Ulcers

Figure 4-15—Change from Baseline to End of Series and ITS Results for Outcome Measures, Pressure Ulcers



Source: The Evaluation Contractor's analysis of HEN-level data.

Among pressure ulcer outcomes, the PfP HENs submitted 161 outcome measures, as shown in Figure 4-15. Of these outcomes, 46.58 percent of the measures (n = 75) were widely and consistently reported. This is the third lowest percentage of widely-reported and consistent measures across AEAs, only slightly higher than falls and ADE. With fewer than half of the measures in the widely-reported and consistent category, the results in this section may not be as representative of the HEN performance as for other AEAs.

Across pressure ulcer measures, 71.43 percent of measures (n = 115) improved by at least 17.6 percent. Beyond the 17.6 percent goal, 39.13 percent of measures (n = 63) achieved a 40 percent reduction. Thus, nearly three-quarters of pressure ulcer outcome measures tested exhibited harm reductions on the scale of the PfP campaign goals.

Of the 161 pressure ulcer outcome measures submitted, 160 measures contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 160 measures tested, 52.50 percent (n = 84) exhibited significant improvement while 8.13 percent (n = 13) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 6.46 to 1, indicating that pressure ulcer outcomes were more than six times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 39.38 percent (n = 63) of the measures exhibited no significant changes from the beginning of reporting to the end.

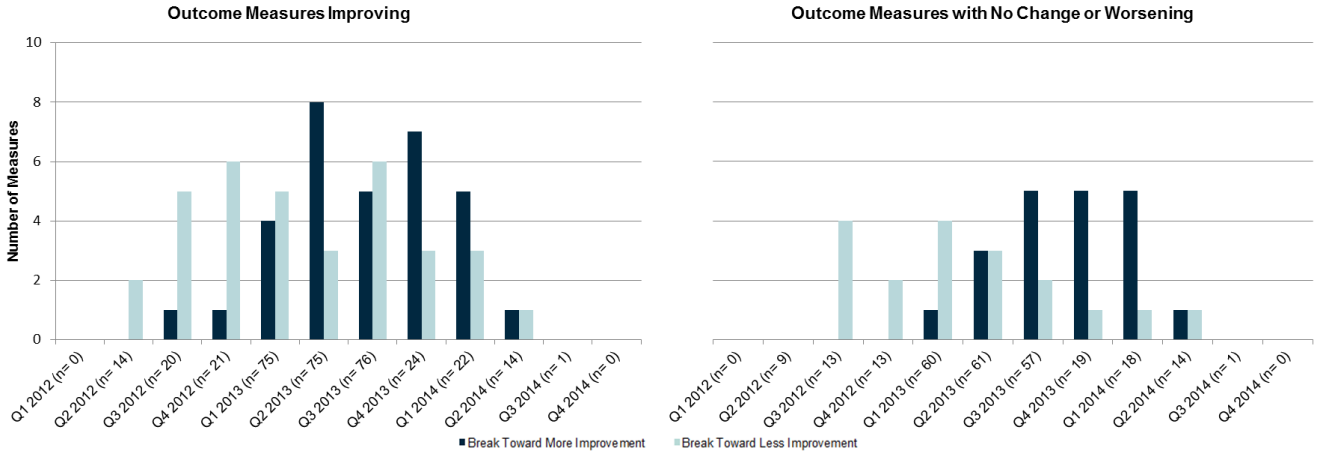
Among the 84 pressure ulcer outcome measures that exhibited significant improvement, the average level of improvement was a 49.48 percent reduction in patient harms, almost three times the required reduction to meet the PfP campaign goals. In contrast, when measures exhibited significant worsening, the average increase in patient harms was 141.96 percent.[4-19] Although measures of patient harms that worsened significantly exhibited rate increases that were substantial, some of these increases may reflect moderate changes in harms with low base rates at baseline.

4-19 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EED from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

Across pressure ulcer outcomes, there were 150 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for pressure ulcer outcomes, 15.33 percent (n = 23) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 10.67 percent (n = 16) exhibited structural breaks associated with changes in trend toward less improvement. Almost three-quarters of the measures (n = 111) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 1,474 tests performed across all measures, with 23.54 percent exhibiting significant structural breaks: 9.29 percent toward greater improvements, and 14.25 toward lesser improvement. The results presented in Figure 4-15, indicate that over 25 percent of the measures changed during the course of the campaign, with 43.67 percent more measures showing trend breaks toward more improvement as those showing trend breaks toward less improvement.[4-20] Thus, these changes in measures are not unexpected and indicate that processes producing pressure ulcer rates were changing toward improvements in the HENs.

Of the 150 measures examined with ITS, 9.33 percent (n = 14) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent 60.86 percent of the 23 measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trends were slightly more likely than not to be associated with overall improvements in measure rates. An additional 44.00 percent (n = 66) of measures still improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-16—Timing of Structural Breaks in Trend for HEN-Level Pressure Ulcer Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.

Figure 4-16 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for pressure ulcers. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-16. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for

4-20 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((15.33– 10.67)/10.67)*100 = 43.67 percent.

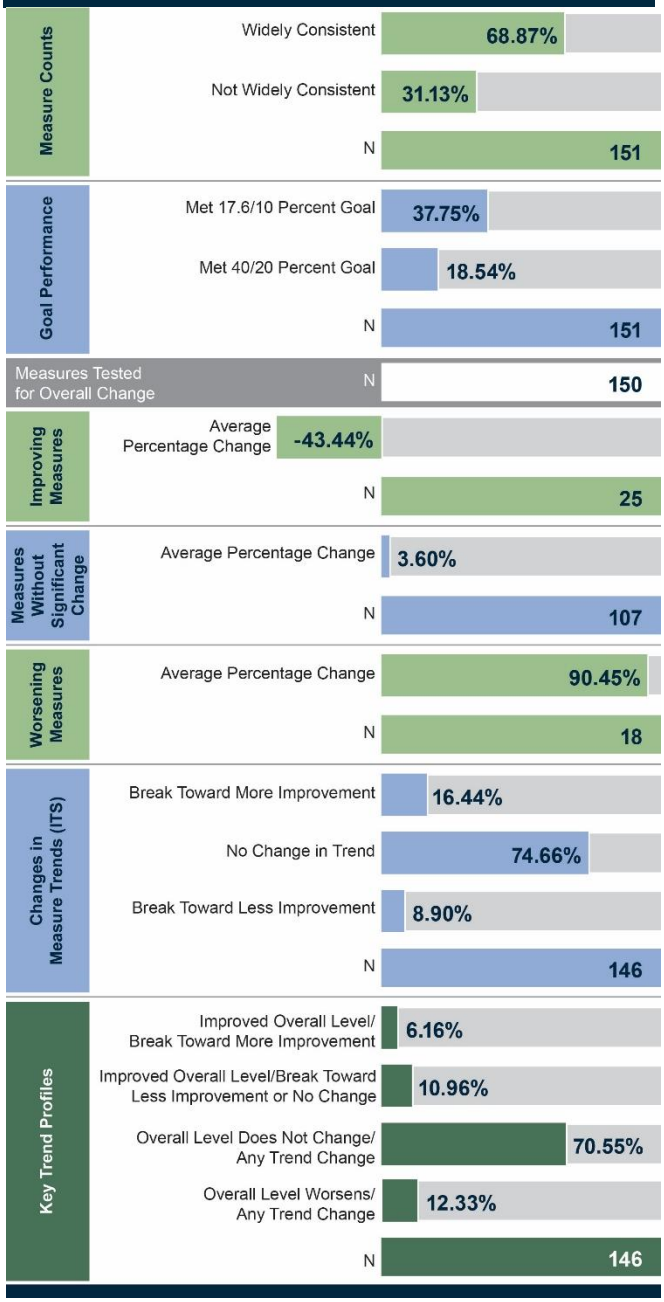
break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are somewhat consistent between measures that improved and those that did not. For both types of measures, the PfP campaign began with a larger proportion of structural breaks toward less improvement. This was followed during the latter half of the campaign by a larger proportion of breaks toward more improvement. This pattern indicates that pressure ulcer outcomes were more likely to exhibit trends of smaller declines or even worsening, before eventually changing toward greater improvements. It is unclear from these results whether this pattern is associated with unsuccessful activities to reduce patient harms followed by the evolution into more successful strategies, or if the focus on pressure ulcers may have simply increased detection within hospitals, thereby raising rates before substantive declines were realized.

Third, it is important to note that measures exhibiting overall improvement appear to have begun exhibiting structural breaks toward more improvement only slightly earlier than measures that did not exhibit overall improvement. For improved measures, the structural breaks toward more improvement began occurring in Q3 2012, and increased to a peak in Q2 2013. Only slightly later in timing, measures that did not improve overall exhibited initial breaks toward more improvement in Q1 2013, and increase to a peak in Q3 2013. Thus, the timings of the structural breaks toward more improvement were approximately similar, with measures showing overall improvement exhibiting favorable breaks about 1 quarter earlier than other measures.

Results for Surgical Site Infections (SSI)

Figure 4-17—Change from Baseline to End of Series and ITS Results for Outcome Measures, SSI



Source: The Evaluation Contractor's analysis of HEN-level data.

Among SSI outcomes, the PfP HENs submitted 151 outcome measures, as shown in Figure 4-17. Of these outcomes, 68.87 percent of the measures (n = 104) were widely and consistently reported. This is the fourth largest percentage of widely-reported and consistent measures across AEAs. With nearly 7 out of 10 measures in the widely-reported and consistent category, the results in this section may be more representative of the HEN performance than for other AEAs.

Across SSI measures, 37.75 percent of measures (n = 57) improved by at least 17.6 percent. Beyond the 17.6 percent goal, 18.54 percent of measures (n = 28) achieved a 40 percent reduction. Thus, just over one-third of SSI outcome measures tested exhibited harm reductions on the scale of the PfP campaign goals.

Of the 151 SSI outcomes submitted, 150 measures contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 150 measures tested, 16.67 percent (n = 25) exhibited significant improvement while 12.00 percent (n = 18) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 1.39 to 1, indicating that SSI outcomes were only 40 percent more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 71.33 percent (n = 107) of the measures exhibited no significant changes from the beginning of reporting to the end, suggesting fewer measures responding substantively to quality improvement efforts.

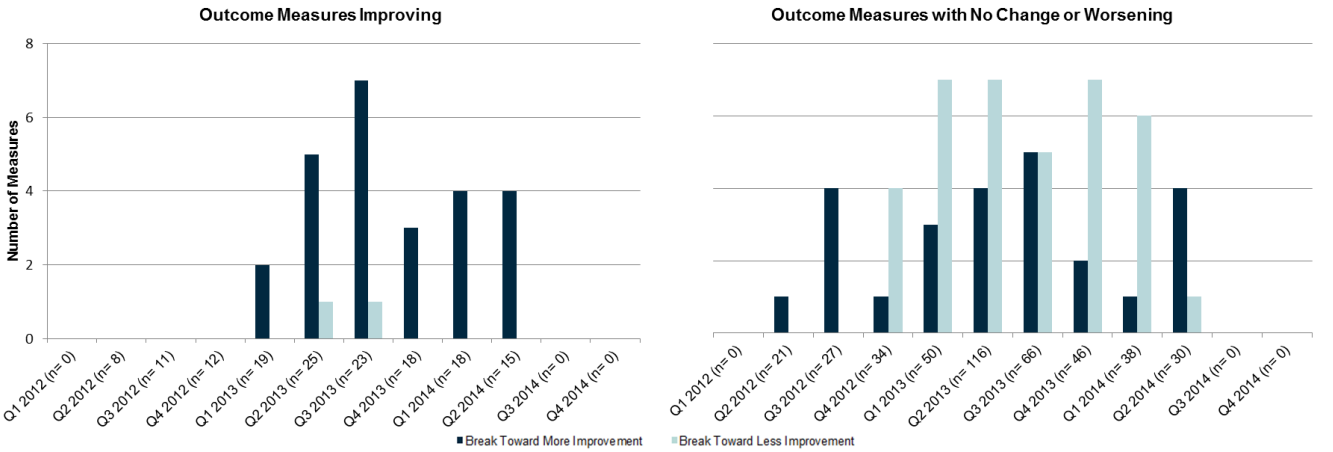
Among the 25 SSI outcome measures that exhibited significant improvement, the average level of improvement was a 43.44 percent reduction in patient harms, more than two times the required reduction to meet the PfP campaign goals. In contrast, when measures exhibited significant worsening, the average increase in patient harms was 90.45 percent.[4-21] Thus, measures of patient harms that worsened significantly exhibited relatively higher rate increases, although some of these increases may reflect moderate changes in harms with low base rates at baseline.

4-21 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EED from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

Across SSI outcomes, there were 146 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for SSI outcomes, 16.44 percent (n = 24) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 8.90 percent (n = 13) exhibited structural breaks associated with changes in trend toward less improvement. Three-quarters of the measures (n = 109) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 2,618 tests performed across all measures, with 14.51 percent exhibiting significant structural breaks: 6.61 percent toward greater improvements, and 7.91 toward lesser improvement. The results presented in Figure 4-17, indicate that approximately one-quarter of the measures changed during the course of the campaign, with 85 percent more measures showing trend breaks toward more improvement as those showing trend breaks toward less improvement.[4-22] Thus, these changes in measure rates are not unexpected and indicate that processes producing SSI rates were changing toward improvements in the HENs.

Of the 146 measures examined with ITS, only 6.16 percent (n = 9) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent about one-third of the 24 measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were less likely than not to be associated with overall improvements in measure rates. Another 10.96 percent (n = 16) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-18—Timing of Structural Breaks in Trend for HEN-Level SSI Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.

Figure 4-18 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for SSI. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-18. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

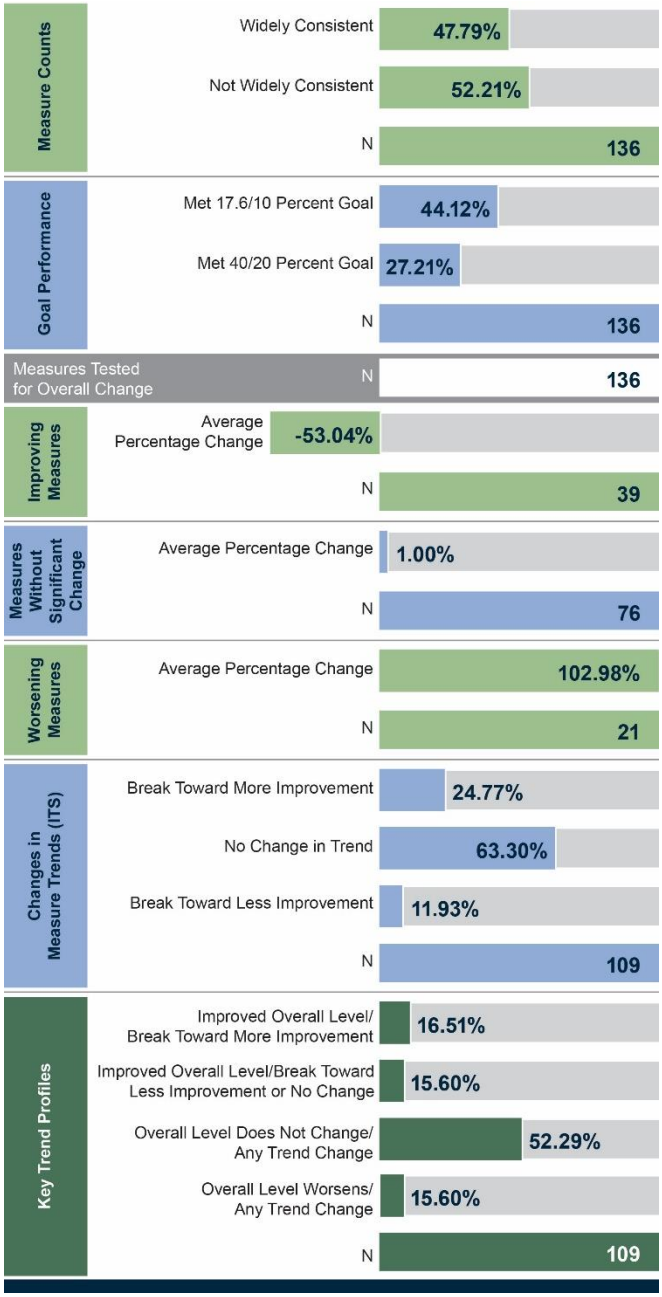
4-22 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((16.44– 8.90)/8.90)*100 = 84.72 percent.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are not consistent between measures that improved and those that did not. SSI outcomes that showed overall improvement were substantially more likely to exhibit structural breaks toward more improvement than toward less improvement. Among the improved measures with significant structural breaks identified in the ITS model, all but two of the breaks were toward more improvement. In contrast, among SSI measures that did not have overall improvement, the predominant structural break identified by the ITS model was toward less improvement, although the structural breaks identified were substantially more mixed among these measures.

Third, it is important to note that measures exhibiting overall improvement appear to have begun exhibiting structural breaks toward more improvement later than measures that did not exhibit overall improvement. For improved measures, the structural breaks toward more improvement began occurring in Q1 2013, and increased to a peak in Q3 2013. In comparison, measures that did not exhibit overall improvements experienced structural breaks toward more improvement as early as Q2 2012, with breaks toward less improvement increasing in Q4 2012. Thus, the timing of the structural breaks was quite different across measures improving versus not improving. The pattern is consistent with an AEA for which significant improvements in outcomes did not occur early during the PfP campaign, and were not widespread across HENs.

Results for Ventilator-Associated Pneumonia (VAP)/Ventilator-Associated Events (VAE)

Figure 4-19—Change from Baseline to End of Series and ITS Results for Outcome Measures, VAP/VAE



Source: The Evaluation Contractor's analysis of HEN-level data.

Among VAP/VAE outcomes, the PfP HENs submitted 136 outcome measures, as shown in Figure 4-19. Of these outcomes, 47.79 percent of the measures (n = 65) were widely and consistently reported.

Across VAP/VAE measures, 44.12 percent of measures (n = 60) improved by at least 17.6 percent. Beyond the 17.6 percent goal, 27.21 percent of measures (n = 37) achieved a 40 percent reduction. Thus, almost half of VAP/VAE outcome measures tested exhibited harm reductions on the scale of the PfP campaign goals.

All of the 136 VAP/VAE measures submitted contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 136 measures tested, 28.68 percent (n = 39) exhibited significant improvement while 15.44 percent (n = 21) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 1.86 to 1, indicating that VAP/VAE outcomes were just under twice as likely to improve during the PfP campaign as to worsen. In contrast to these changing measures, 55.88 percent (n = 76) of the measures exhibited no significant changes from the beginning of reporting to the end, suggesting about half of the measures responded substantively to quality improvement efforts.

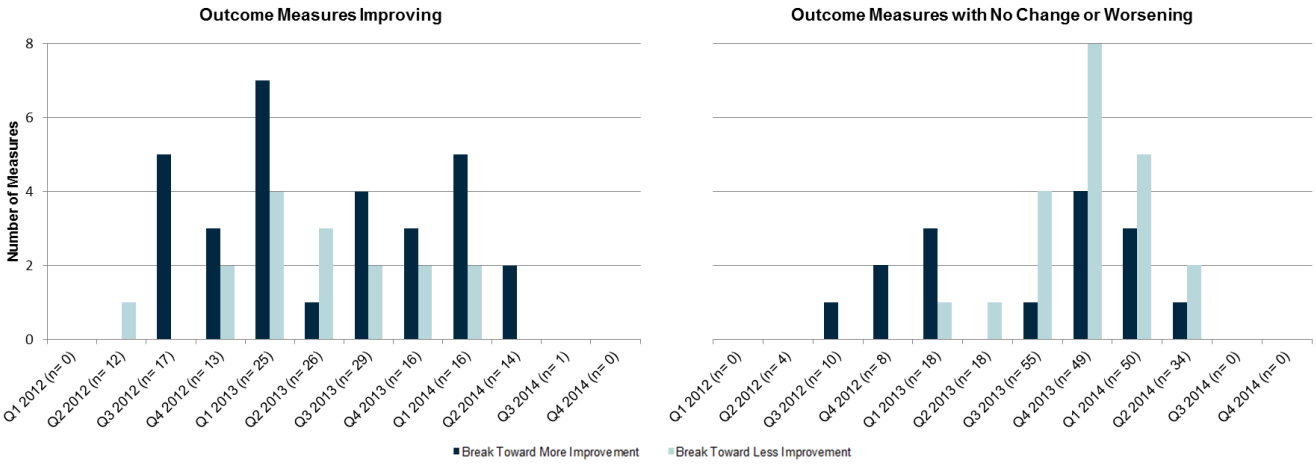
Among the 39 VAP/VAE outcome measures that exhibited significant improvement, the average level of improvement was a 53.04 percent reduction in patient harms, more than three times the required reduction to meet the PfP campaign goals. In contrast, when measures exhibited significant worsening, the average increase in patient harms was 102.98 percent.[4-23] Although measures of patient harms that worsened significantly exhibited rate increases that were substantial, some of these increases may reflect moderate changes in harms with low base rates at baseline.

4-23 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EED from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

Across VAP/VAE outcomes, there were 109 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for VAP/VAE outcomes, 24.77 percent (n = 27) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 11.93 percent (n = 13) exhibited structural breaks associated with changes in trend toward less improvement. Over 60 percent of the measures (n = 69) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 1,386 tests performed across all measures, with 18.61 percent exhibiting significant structural breaks: 10.39 percent toward greater improvements, and 8.23 toward lesser improvement. The results presented in Figure 4-19, indicate that more than one-third of the measures changed during the course of the campaign, with more than twice as many measures showing trend breaks toward more improvement as those showing trend breaks toward less improvement.[4-24] Thus, these changes in measure rates are not unexpected and indicate that processes producing VAP/VAE rates were changing toward improvements in the HENs.

Of the 109 measures examined with ITS, only 16.51 percent (n = 18) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent two-thirds of the 27 measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were more likely than not to be associated with overall improvements in measure rates. Another 15.60 percent (n = 17) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-20—Timing of Structural Breaks in Trend for HEN-Level VAP/VAE Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.

Figure 4-20 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for VAP/VAE. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-20. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

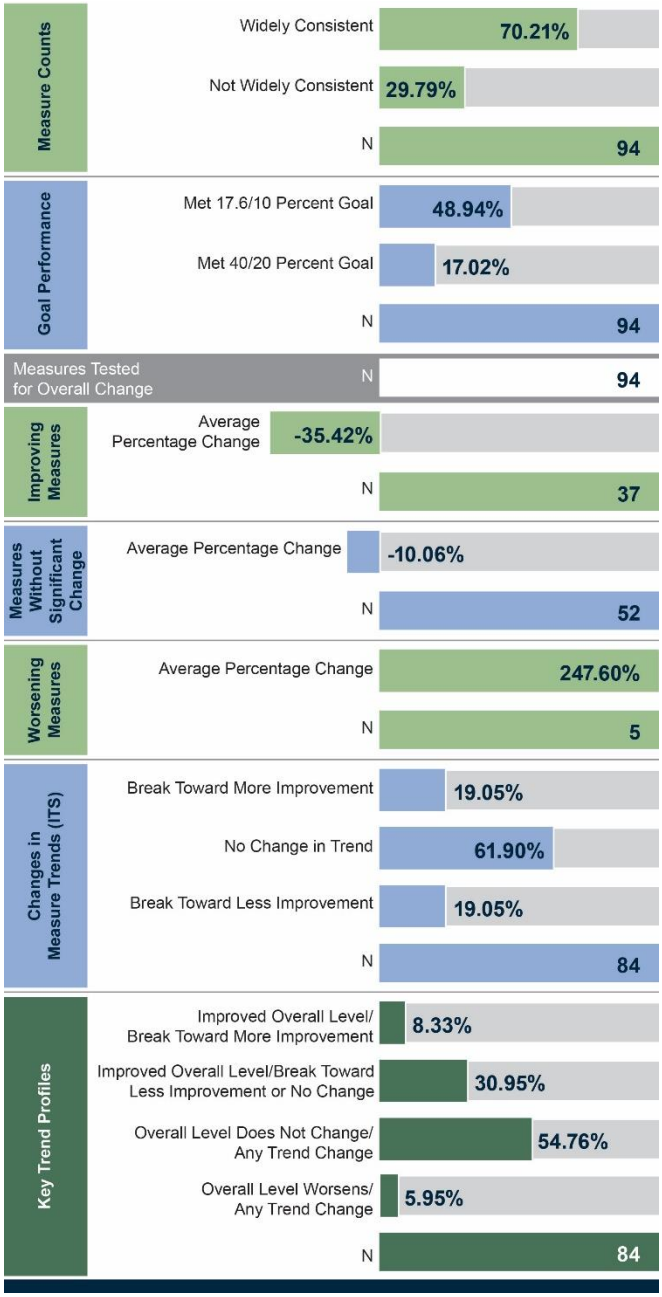
4-24 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((24.77– 11.93)/11.93)*100 = 107.63 percent.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are not consistent between measures that improved and those that did not. VAP/VAE outcomes that showed overall improvement were substantially more likely to exhibit structural breaks toward more improvement than toward less improvement. This trend held throughout the PfP campaign period. In contrast, the VAP/VAE outcomes that did not show overall improvement were more likely to begin the campaign with some structural breaks toward more improvement through Q1 2013. However, beginning in Q1 2013, structural breaks toward less improvement began appearing, and increased to peak in Q4 2013 outpacing more favorable structural breaks.

Third, it is important to note that measures exhibiting overall improvements were more likely to exhibit structural breaks toward more improvement throughout the campaign than measures that did not exhibit overall improvement. In contrast, measures that did not exhibit overall improvement were more likely to exhibit structural breaks toward less improvement during the latter half of 2013 and first half of 2014. Thus, the timing of the structural breaks was quite different across measures improving versus not improving. The pattern likely reflects, at least in part, a substantive change in measure definitions that occurred in the second half of 2013 as hospitals and HENs switched from reporting VAP to several different measures of VAE, probable/possible VAP, and infection-related ventilator-associated complications (IVAC). The change in definition was motivated both by an industry-wide understanding that previously measures of VAP were limited in scope, and by a change made by the Centers for Disease Control and Prevention's (CDC's) National Healthcare Safety Network (NHSN) to collect data for a variety of ventilator-associated harm measures. The CDC's NHSN transition occurred in January 2013. As more hospitals began reporting the new measures, the data indicates that structural breaks toward less improvement were more likely to occur. This is consistent with HENs and hospitals reporting on new measures for which established baselines were not available, and observing an increasing rate as monitoring and detection improved.

Results for Venous Thromboembolism (VTE)

Figure 4-21—Change from Baseline to End of Series and ITS Results for Outcome Measures, VTE



Source: The Evaluation Contractor's analysis of HEN-level data.

Among VTE outcomes, the PfP HENs submitted 94 outcome measures, as shown in Figure 4-21. Of these outcomes, 70.21 percent of the measures (n = 66) were widely and consistently reported. This is the third largest percentage of widely-reported and consistent measures across all AEAs, increasing confidence that the results observed here are more likely to be representative of the performance of HEN-aligned hospitals.

Across VTE measures, 48.94 percent of measures (n = 46) improved by at least 17.6 percent. Beyond the 17.6 percent goal, 17.02 percent of measures (n = 16) achieved a 40 percent reduction. Thus, approximately half of VTE outcome measures tested exhibited harm reductions on the scale of the PfP campaign goals.

All of the 94 VTE measures submitted contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 94 measures tested, 39.36 percent (n = 37) exhibited significant improvement while 5.32 percent (n = 5) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 7.40 to 1, indicating that VTE outcomes were seven times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, 55.32 percent (n = 52) of the measures exhibited no significant changes from the beginning of reporting to the end, suggesting that slightly under half of the measures responded substantively to quality improvement efforts.

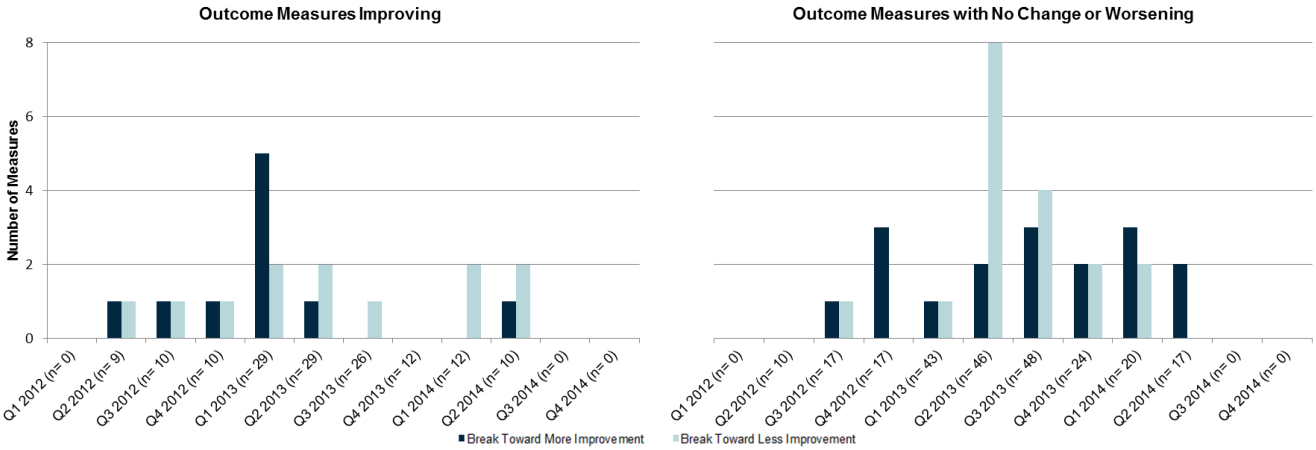
Among the 37 VTE outcome measures that exhibited significant improvement, the average level of improvement was a 35.42 percent reduction in patient harms, more than twice the required reduction to meet the PfP campaign goals. In contrast, when measures exhibited significant worsening, the average increase in patient harms was 247.60 percent.[4-25] Thus, measures of patient harms that worsened significantly exhibited rate increases that were substantial, although some of these increases may reflect moderate changes in harms with low base rates at baseline.

4-25 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EED from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

Across VTE outcomes, there were 84 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for VTE outcomes, 19.05 percent (n = 16) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 19.05 percent (n = 16) exhibited structural breaks associated with changes in trend toward less improvement. Over 60 percent of the measures (n = 52) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 1,405 tests performed across all measures, with 24.84 percent exhibiting significant structural breaks: 12.53 percent toward greater improvements, and 12.31 toward lesser improvement. The results presented in Figure 4-21, indicate that more than one-third of the measures changed during the course of the campaign, with the same proportion of measures showing trend breaks toward more improvement as those showing trend breaks toward less improvement.[4-26] Thus, these changes in measures are not unexpected and indicate that processes producing VTE rates favored quality improvements, but were not more likely to exhibit structural breaks indicating greater improvement.

Of the 84 measures examined with ITS, only 8.33 percent (n = 7) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represent almost half of the 16 measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were slightly less likely than not to be associated with overall improvements in measure rates. Another 30.95 percent (n = 26) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-22—Timing of Structural Breaks in Trend for HEN-Level VTE Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.

Figure 4-22 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for VTE. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-22. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

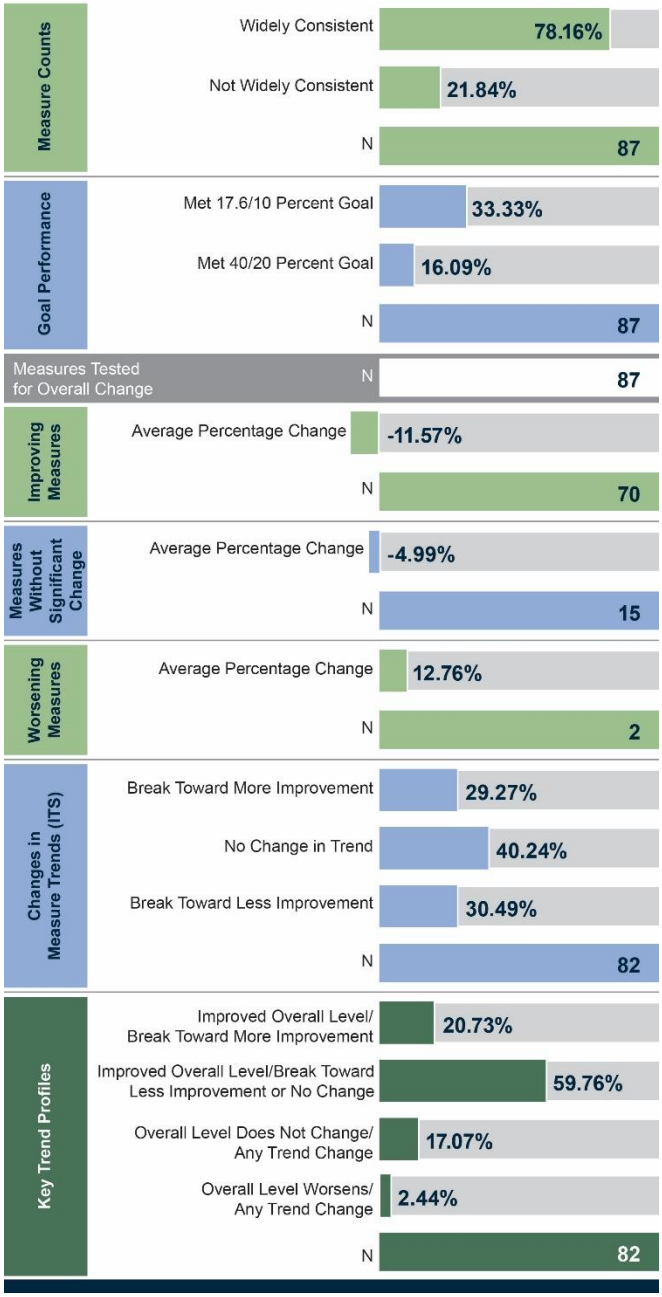
4-26 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((19.05– 19.05)/19.05)*100 = 0 percent difference.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are not consistent between measures that improved and those that did not. VTE outcomes that showed overall improvement were substantially more likely to exhibit structural breaks toward more improvement than toward less improvement in Q1 2013, while measures that did not improve were substantially more likely to exhibit breaks toward less improvement in Q2 2013. Both groups of measures do not exhibit many significant breaks, and the latter half of 2013 and first half of 2014 were more likely to exhibit structural breaks toward less improvement.

Third, it is important to note that the maximum number of significant breaks detected by the ITS model in any quarter was 7 for measures that improved, and 10 for measures that did not improve. However, the modal number of breaks across all quarters was four for measures that improved overall, and two for measures that did not improve overall. This pattern does not provide strong evidence that the timing of changes in the HEN-level VTE rates is clearly associated with the overall improvements exhibited in this AEA.

Results for Readmissions

Figure 4-23—Change from Baseline to End of Series and ITS Results for Outcome Measures, Readmissions



Source: The Evaluation Contractor's analysis of HEN-level data.

Among readmission outcomes, the PfP HENs submitted 87 outcome measures, as shown in Figure 4-23. Of these outcomes, 78.16 percent of the measures (n = 68) were widely and consistently reported. This is the largest percentage of widely-reported and consistent measures across all AEAs, increasing confidence that the results observed here are more likely to be representative of the performance of HEN-aligned hospitals.

Across readmission measures, 33.33 percent of measures (n = 29) improved by at least 10 percent. Beyond the 10 percent goal, 16.09 percent of measures (n = 14) achieved a 20 percent reduction. Thus, approximately one third of readmission outcome measures tested exhibited harm reductions on the scale of the PfP campaign goals.

All of the 87 readmissions measures submitted contained sufficient information to determine whether the measures exhibited a statistically significant change. Out of the 87 measures tested, 80.46 percent (n = 70) exhibited significant improvement while 2.30 percent (n = 2) exhibited significant worsening. Therefore, the ratio of measures improving to measures worsening is 34.98 to 1, indicating that readmission outcomes were almost 35 times more likely to improve during the PfP campaign than to worsen. In contrast to these changing measures, only 17.24 percent (n = 15) of the measures exhibited no significant changes from the beginning of reporting to the end, suggesting about half of the measures responded substantively to quality improvement efforts.

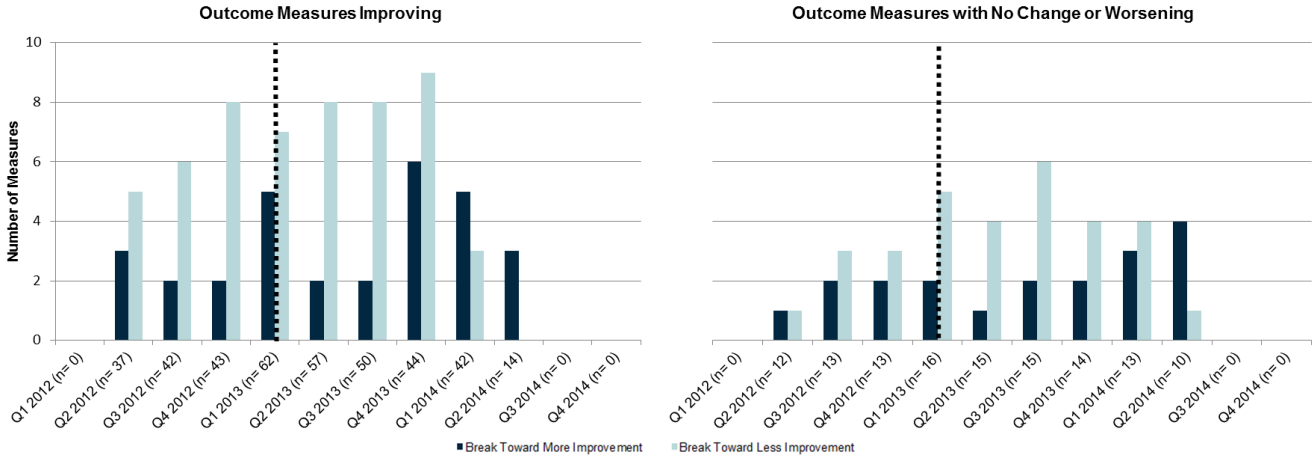
Among the 70 readmission outcome measures that exhibited significant improvement, the average level of improvement was an 11.57 percent reduction in patient harms, more than the required reduction to meet the PfP campaign goals. In contrast, when measures exhibited significant worsening, the average increase in patient harms was only 12.76 percent.[4-27] Thus, measures of patient harms that worsened significantly exhibited rate increases that were nontrivial, although some of these increases may reflect moderate changes in harms with low base rates at baseline.

4-27 The average increase in patient harms that worsened is larger than the average reduction in patient harms that were reduced in part because measures with very low base rates may exhibit small absolute changes that result in large percentage differences. For example, a HEN with an OB-EED rate of 0.5 percent at baseline is very low in comparison to other HENs. If that HEN experienced an increase in OB-EED from 0.5 to 1.5, its rate would still be very low, but would have exhibited ((1.5 – 0.5) / 0.5)*100 = 200 percent increase.

Across readmissions outcomes, there were 82 measures with at least eight data points, the minimum threshold used in this report for ITS analysis. Among the ITS models estimated for readmissions outcomes, 29.27 percent (n = 24) of the measures exhibited structural break points associated with a change in trend toward greater improvement. Another 30.49 percent (n = 25) exhibited structural breaks associated with changes in trend toward less improvement. Over 40 percent of the measures (n = 33) exhibited no significant structural break. Given such a large number of statistical tests, significant results would be expected to occur by chance for about 5 percent of the tests performed: 2.5 percent improving, and 2.5 percent worsening. There were 1,685 tests performed across all measures, with 23.92 percent exhibiting significant structural breaks: 9.08 percent toward greater improvements, and 14.84 toward lesser improvement. The results presented in Figure 4-23, indicate that almost two-thirds of the measures changed during the course of the campaign, with the same proportion of measures showing trend breaks toward more improvement as those showing trend breaks toward less improvement.[4-28] Thus, these changes in measures are not unexpected and indicate that processes producing readmission rates much more likely to exhibit improvements, but the structural breaks in readmissions were split between breaks toward more improvement and those toward less improvement.

Of the 82 measures examined with ITS, 20.73 percent (n = 17) exhibited both an overall improvement in rate from baseline to the most recent 3 months reported, and a structural break in trend toward greater improvements during the PfP campaign. These represented 70 percent of the 24 measures with structural breaks toward more improvement, indicating that quality improvements resulting in significant improvement in trend were more likely than not to be associated with overall improvements in measure rates. Another 59.76 percent (n = 49) also improved, but either without a structural break in trend, or a break toward less improvement.

Figure 4-24—Timing of Structural Breaks in Trend for HEN-Submitted Readmissions Outcomes



Source: The Evaluation Contractor's analysis of HEN-level data.
Note: The dashed line in the figure represents the following campaign-level event: HENs readmissions commitments submitted to CMS (January 2013).

Figure 4-24 shows two graphs that illustrate the timing of structural breaks in HEN outcomes for readmissions. The left panel shows the timing of breaks for measures that exhibited overall improvements, while the right panel shows the timing of breaks for measures that did not improve overall. There are three important results to note in Figure 4-24. First, because the ITS analysis requires at least four observations in the pre- and post-intervention periods, no breaks could be identified during Q1 2012 or Q4 2014, the first and

4-28 The relative difference in measures exhibiting breaks toward more improvement versus measures exhibiting break toward less improvement is calculated as ((30.49-29.27)/30.49)*100 = 4.00 percent.

last complete quarters of HEN activity, respectively. As a corollary to this point, more measures are tested for break points during the middle portion of the campaign than at the beginning or end, and the bell-shaped curve shown in the graph is largely due to the change in the number of measures being tested in each quarter.

Second, the ratio of breaks toward more improvement and breaks toward less improvement are somewhat consistent between measures that improved and those that did not. Both types of outcomes for readmissions were substantially more likely to exhibit structural breaks toward less improvement than toward more improvement for the majority of the campaign. For both groups of measures, the number of structural breaks identified by ITS increased slightly over time, peaking in Q3 or Q4 2013. However, the number of breaks identified for measures that exhibited improvement was greater than for measures without improvement.

Third, it is important to note that the national trend in Medicare fee-for-service (FFS) 30-day all-cause readmissions during the PfP campaign was a downward trend beginning in mid-2011, and then leveling off in Q1 2014. The large number of structural breaks toward less improvement is likely a reflection of HEN readmission rates dropping and then leveling off at different points throughout the campaign. For measures that improved overall, as well as those that did not, the number of breaks toward less improvement is reduced in Q1 and Q2 of 2014, consistent with national trends.

Conclusion

The results presented above describe a campaign for which outcomes associated with patient safety improved substantially, if in varied amounts across topic areas. The AEAs for OB-EED, pressure ulcers, and VTE had the largest percentages of measures reach the 17.6 percent goal, and OB-EED and pressure ulcers had two of the largest three reductions in patient harms. In contrast, CAUTI, OB-Other, and readmissions had the smallest percentages of measures achieve the17.6 percent/10 percent goal, and CAUTI and OB-Other also had relatively smaller reductions in patient harms.[4-29]

Across all AEAs, the proportion of measures that exhibited improvement was substantively larger than the proportion of measures that worsened. However, for ADE (76/60), falls (67/51), OB-EED (30/7), pressure ulcers (84/63), and readmissions (70/15), the proportion of measures exhibiting reductions in patient harm exceeded the proportion of measures exhibiting no change in patient harms. Combining the ratio of improving measures to worsening measures, and the relative number of measures improving to those exhibiting no change, the data indicate that the measures for CLABSI, falls, OB-EED, and readmissions were the most likely to exhibit improvement across all AEAs.

With respect to the magnitude of improvements, OB-EED, VAP/VAE, and pressure ulcers exhibited the largest average improvements. Although the goals were measured on different scales, readmissions also exhibited substantial average improvements. In contrast, CAUTI, CLABSI, and falls exhibited the smallest average improvements.

The results from the ITS analysis indicates that for many AEAs, less than one-third of the measures exhibited significant structural breaks toward greater improvements during the period of the PfP campaign. In comparison, the percentage of measures worsening was more than would be expected by random chance, but was slightly less than the percentage improving. The overall results suggest that there were slightly more structural breaks of a favorable nature. CAUTI, OB-EED, readmissions, and ADE exhibited the largest percentages of measures with structural breaks toward more improvement. Pressure ulcers, SSI, VTE, and CLABSI exhibited the fewest structural breaks toward more improvement.

[4-29] The average reduction in readmissions among measures that improved was 11.57 percent among 70 out of 85 measures.

When combining the overall improvement results, with the ITS analysis, OB-EED and readmissions stand out as the AEAs for which more than 20 percent of the measures demonstrated overall improvement and structural breaks toward more improvement. These results provide consistent evidence of successful quality improvement activities occurring during the time of the PfP campaign. Importantly, however, these results do not speak to the activities causing the improvements observed in the HEN-level data.

Repeated Measures Analysis of the Association between HEN Activities and Partnerships and Common Measure Outcomes

Consistent with the national trends in patient harms, the HEN-level data exhibit significant and substantial reductions in patient harms during the course of the PfP campaign. The analyses presented earlier in this chapter demonstrate that 616 out of 1,450 (42.48 percent) outcome measures exhibited significant reductions in patient harms. Of the 1,265 measures that could be tested using ITS analysis, 149 measures (11.78 percent) exhibited both an overall improvement in rate and a structural break indicating an acceleration of improvement during the PfP campaign. Having observed these results, the evaluation now turns to assessing the extent to which the activities engaged in and partnerships formed by the HENs were associated with the reductions in patient harms.

To examine the association between HEN activities and partnerships and changes in outcome measures over time, an ideal analysis would make use of the time series aspects of the data and assess the relationship between changes in the number of activities or partnerships and changes in the level of the outcome. This analysis would allow the Evaluation Contractor to determine whether increasing levels of specific activities or types of partnerships are associated with reductions in outcome measures of patient harms. However, due to the limited number of observations for any single measure and individual HEN, the Evaluation Contractor uses a repeated measures mixed model to pool the data across HENs that submitted data on similar measures, also known as common measures. There are 37 common measures, obtained from five different data sources, resulting in 50 different models associating HEN activities and partnerships with outcomes.[4-30]

The repeated measures mixed model was estimated once for each of the common measure and data source combinations. The model pools together the data series for each HEN submitting the relevant measure to increase the statistical power of the analysis, and estimates a unique trend for each HEN that submitted data on a specific common measure, as well as a HEN-specific starting level in order to capture the unobserved differences between HENs and over time periods. The mixed model also identifies the average relationship between activities, partnerships and outcomes. Additional details on the repeated measures mixed model methodology are found in Appendix D.

To the extent that the HEN activities and partnerships are associated with reductions in patient harms, the expected result is that the mixed model will identify a negative relationship over time between these variables. Table 4-1 provides a summary of the results found using the mixed models.

Each of the four types of HEN activities and eight types of partnerships were tested for a relationship over time in 50 different models.[4-31] Two additional covariates – a repeated measure dummy variable for quarters and the quarterly number of hospitals reporting – were also included as control variables. This results in the potential for 700 coefficient estimates. However, for 86 of the coefficients estimated in the models, the covariate did not exhibit sufficient unique variation to estimate an independent association with the outcome.

4-30 A full list of the 37 common measures can be found in Appendix E.

4-31 One measure and data source combination did not have sufficient variation for the model to be estimated. Therefore, the total number of common measure by data source combinations is 49.

Therefore, only 614 coefficients were estimated in the mixed models. Of the estimated coefficients, only 79 (12.87 percent) exhibited statistically significant relationships with the patient harm outcomes modeled. Tool dissemination and one-on-one coaching, along with federal partnerships exhibited the largest number of significant relationships, at 10, 9, and 8 respectively. Among the 79 significant relationships observed, there was nearly an even divide between the 38 (48.10 percent) with the expected negative relationship, and the 41 (51.90 percent) with an unexpected positive relationship.

Table 4-1—Statistically Significant Findings From Repeated Measures Mixed Models Analyses

Covariate Group	Covariate	Coefficients not Estimated (Insufficient Variation)	Coefficients Estimated	Significant Coefficients	Covariate Associated with Harm Reduction (Negative Coefficient)	Covariate Associated with Harm Increase (Positive Coefficient)
Initiatives	Tools Initiatives	2	48	10	4	6
	Education Initiatives	1	49	5	1	4
	Coaching Initiatives	2	48	9	5	4
	Leadership Initiatives	3	47	2	2	0
Partners	Federal Partners	11	39	8	4	4
	National Private Partners	6	44	3	1	2
	State and Local Health Organization Partners	9	41	6	2	4
	State and Local Private Organization Partners	2	48	6	2	4
	State Hospital Association Partners	18	32	4	4	0
	Subject Matter Expert Partners	10	40	5	3	2
	Other HEN Partners	7	43	3	1	2
	Other Partners	13	37	7	1	6
Controls	Quarter Number	1	49	5	4	1
	Engaged Hospitals	1	49	6	4	2
Totals		**86**	**614**	**79**	**38**	**41**

Source: Evaluation Contractor's analysis of HEN-submitted data, November 2014 and HEN timeline data, fall 2014.
Notes: Statistical significance reached at $p \leq 0.0$. There are 773 reported measures, 742 of which had usable data. There are 50 combinations of HENS, cohorts, and data sources produced 49 measures with sufficient variation to model. *t*-test shows no difference between the numbers of negative and positive coefficients ($p = .675$)

The division of activity and partnership coefficients between negative and positive associations suggests that the directions of association are largely random in nature. Only one activity (i.e., leadership initiatives) and one partnership type (i.e., state hospital association partners) exhibited coefficients that were all in the expected negative direction. Increases in leadership activities were associated with reductions in CLABSI per 1,000 central line days in both HEN-sourced and National Database of Nursing Quality Indicators® (NDNQI®) data.[4-32] Increases in state hospital association partners lead to reductions in the HEN-submitted CAUTI Standardized Infection Ratio (SIR), and three models for pressure ulcers: NDNQI all-stage hospital acquired pressure ulcers per 100 assessed patients, NDNQI stage 2+ hospital acquired pressure ulcers per

4-32 NDNQI® is a registered trademark of the American Nurses Association (ANA). NDNQI® data were supplied by ANA. The ANA disclaims responsibility for any analyses, interpretations, or conclusions.

100 assessed patients, and HEN-submitted Agency for Healthcare Quality Research (AHRQ) Patient Safety Indicator (PSI)-03 stage 3+ pressure ulcers.[4-33] The detailed results for these models can be found in Appendix E.

In contrast to the limited but consistently negative association between leadership activities and state hospital association, the analysis found that all other HEN activities and partnership types exhibited mixed results, being associated with reductions in patient harms for some models, and increases in patient harms for other models. Therefore, the results do not point toward a consistently negative association between these HEN characteristics and patient harms. Rather, the result suggests that the observed relationships are likely to be randomly distributed between positive and negative associations.

One reason why this randomness in the results may occur is that the HEN-reported activity and partnership data are less reliable than would be preferred. Specifically, the HENs were asked to systematically report the activities they engaged in and partnerships that they formed during the last full month of the PfP campaign. Prior to that period, these data were reported in varying forms and with varying frequency in the HEN monthly reports. The retrospective nature of the systematic data collection effort is a likely source of unreliability in the HEN activity and partnership data. The HEN staff may not have been able to recall all of the activities, and precise timing of activities, over the entire three year period. Furthermore, detailed examination of the activity data indicated that some HENs may have reported a more comprehensive set of activities and others a more limited set. The result this reporting by the HENs would be the weakening of any discernable pattern of relationships.

[4-33] The definitions for the AHRQ PSIs can be found here: http://www.qualityindicators.ahrq.gov/modules/PSI_TechSpec.aspx.

Analyses with Comparison Groups

The Evaluation Contractor defined the intervention group as HEN-aligned hospitals – those hospitals that signed an agreement to align with a HEN by June 2012. By aligning with the HEN, the hospital was committing to reducing harms, and would receive the services that the HEN provided as described in Chapter 3. Hospitals' alignment with a HEN is based on monthly rosters submitted by HENs. Any hospital that signed an agreement to align with the HEN between July 2012 and December 2012 was excluded from the analysis, as they received some services, but did not receive the "full dose." All short-term acute care hospitals in the nation that are not on the HENs' roster are considered not to be HEN-aligned and, are used as comparison hospitals in all difference-in-differences analyses presented in this report. The actual hospitals used in each analysis varies by the data source, and Appendix D shows the characteristics for the HEN-aligned and non-HEN-aligned.

The effectiveness of the HEN impact was assessed using multiple national-level sources of data, with a difference-in-differences approach used to "net out" the other initiatives that were occurring during this period. The ability to detect differences was limited by three critical components of the PfP campaign, in addition to data limitations that vary by data source. First, HENs were so successful at recruiting hospitals into their network (72 percent of the eligible hospitals aligned with a HEN) that it was difficult to identify non-HEN-aligned hospitals to form an appropriate comparison group. Second, because these hospitals voluntarily aligned with the HEN, there was concern that there was "self-selection" – that is that those hospitals that did not need the resources of the HENs to change were less likely to align with a HEN. Third, because HENs sometimes included non-HEN-aligned hospitals in their learning activities some of the non-HEN-aligned hospitals were exposed to the HEN benefit. All three of these issues weaken the probability that the Evaluation Contractor's analyses will detect a difference between the two groups. However, the Evaluation Contractor used propensity score reweighting to produce a comparison group of non-HEN-aligned hospitals that are similar to HEN-aligned hospitals on observable characteristics of hospitals and their patients. The balance that the Evaluation Contractor was able to achieve between the two hospital groups met the standard for propensity score matching (What Works Clearinghouse, 2010).

Summary of Results

Using a Bayesian analysis, the Evaluation Contractor examined whether there was evidence of a difference in results between the HEN group and the comparison group in rates of four patient harms: Medicare FFS 30-day all-cause readmissions (readmissions) and the AHRQPSIs for central venous catheter-related blood stream infection rate (CRBSI) (PSI-07), pressure ulcer rate (PSI-03), and VTE rate (PSI-12). These outcomes were calculated from Medicare claims data. The analysis found there was a moderate probability of slight or small impact for two (VTE and CRBSI, respectively), a moderate probability of substantial impact for one (pressure ulcers), and no impact for one (readmissions) (For purposes of description, high probability is 85 percent or more, moderate is 60 to 85 percent, and low is less than 60 percent probability. Slight impact is 2 to 5 percent, small is 5 to 10 percent, and substantial is 25 percent or more. Less than 2 percent impact is considered no impact.). However, for nearly all sets of estimates, the uncertainty interval around the impact estimate includes zero, so it is possible that the observed improvements are due to chance variation rather than HEN alignment.[4-34] The presentation of those results is followed by the Bayesian analyses of differences in subsets between specific subgroups of hospitals and HENs.

Difference-in-differences analyses were unable to identify evidence of a HEN impact on OB-EED, birth outcomes, or various measures of patient harms obtained from chart-reviewed data.

[4-34] The one exception–readmissions–shows a magnitude of effect so slight that the Evaluation Contractor concluded it was not meaningful and categorized it as "no impact."

Difference-in-Differences Analysis of Adverse Events and Readmissions with Medicare Claims Data

The key research question for the evaluation is, "is HEN-alignment associated with improved patient safety outcomes overall or for subgroups of hospitals?" The Evaluation Contractor approached this question by applying a Bayesian impact estimation model, which permits intuitive, flexible inferential statements and heightens precision when examining subgroup effects.

The adverse event outcomes of interest in this evaluation not only occur rarely, but also are very rarely recorded in administrative data, resulting in very low statistical power to estimate HEN alignment's impact using conventional (frequentist) methods. As a result, even large impact estimates of the HEN component cannot be deemed "statistically significant" using the frequentist approach, as shown in Chen et al. 2014. Assessing the impact of HEN alignment probabilistically, as Bayesian techniques permit, maintains a rigorous statistical standard while providing a more flexible interpretation of the program's effects. Whereas the frequentist approach classifies the HEN component's impact on a given outcome as significant or not significant, a Bayesian analysis facilitates more granular inference of the form "there is a 46 percent chance that HEN alignment reduced the VTE rate (PSI-12) by 5 percent or more in the post-intervention period (2012–2013)," as the impact estimates reported in the following section indicate. Such conclusions offer the opportunity to tailor inference to substantive questions of interest and to apply subject-matter expertise in deeming effects meaningful.

The Bayesian approach also facilitates subgroup analyses. Whereas the smaller sample sizes available for subgroup analyses tend to decrease power and precision in a traditional framework, Bayesian methods retain precision, enabling the Evaluation Contractor to examine variation in the effects of HEN alignment across subgroups defined based on hospital and HEN characteristics, key questions of interest to policymakers.

As described in Chapter 1, the Evaluation Contractor estimated a Bayesian model and computed impacts at follow-up for 2012, 2013, and for the 2 years combined. Because the most recent data available included the first quarter of 2014 for the analysis of impacts on adverse events, year 2013 includes all four quarters of 2013 and the first quarter of 2014. For the analysis of impacts on readmissions, 2013 includes outcomes only for that year, since the first quarter of 2014 was needed to assess whether readmissions occurred after an index admission at the end of 2013. In both figures and tables, the impacts for 2012 and 2013 are presented separately, but the discussions focus on the impacts in both years combined. Like the frequentist approach, the Bayesian analysis uses a comparison group difference-in-differences design to identify effects attributable to HEN activities. Although the frequentist and Bayesian models take similar forms, the Bayesian model assumes that model parameters have prior probability distributions, which enable probabilistic inference. For more details on the technical specification of the model and the two-stage modeling approach, see Appendix D.

In this chapter, the Evaluation Contractor presents the results of the more traditional frequentist impact estimation model as context for the Bayesian analytical results. Even though the frequentist and Bayesian models differ in many ways, they generally produce comparable results and similar conclusions about the effect of HEN alignment on patient outcomes. The reason for any differences in the results is that the Bayesian model exploits the variation in hospital subgroups, whereas the frequentist model controls for all hospital-specific effects. For more detailed treatments of both approaches, see Appendix D.

After briefly describing the data and the sample, this section opens with a description of Bayesian estimates' probabilistic interpretation. The first subsection describes the results for VTE while introducing key concepts in Bayesian inference that enable probabilistic interpretation. The next subsection then presents impact estimates for pressure ulcer stages 3 and higher (PSI-03) rates, CRBSI rates (PSI-07), and 30-day all-cause readmissions.

Data

The analysis used Medicare claims data for acute care hospitals (except pediatric and psychiatric) that consistently reported present-on-admission indicators, and for which data on hospital characteristics were available for propensity score matching. For these hospitals, all discharges for Medicare FFS beneficiaries were included in the analysis. Additional detail about Medicare data used for this and other analyses is presented in Appendix B. More detail about the sample and about the outcome construction for the Medicare analysis is found in Appendix D. Propensity score reweighting is further described in Appendix D.

Interpreting Bayesian Results

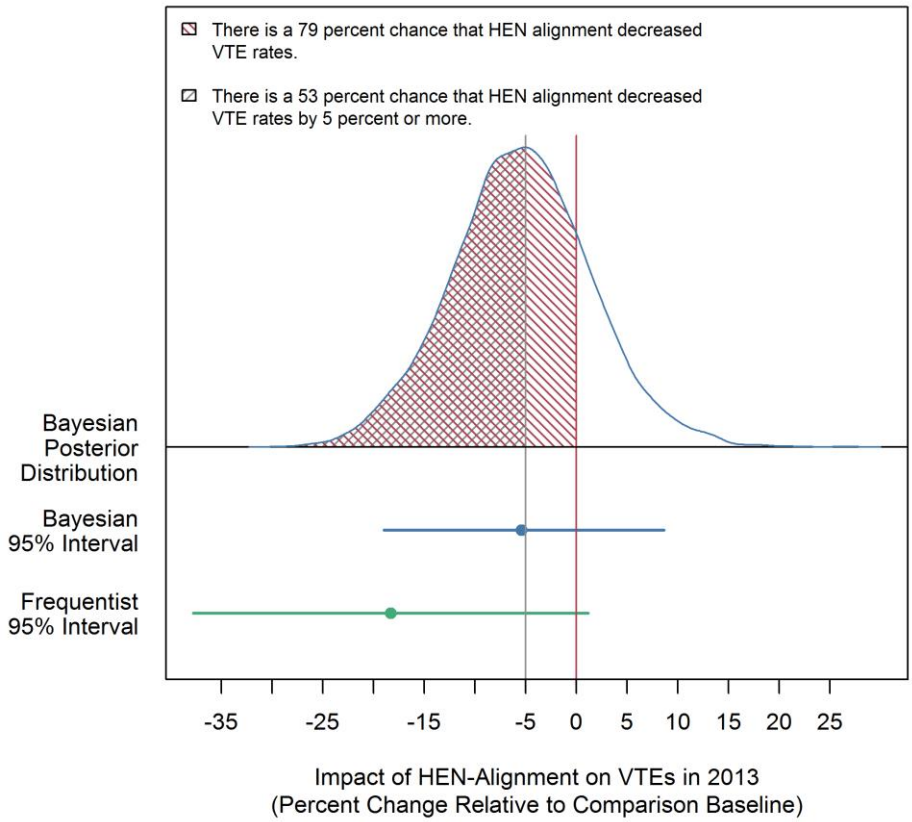
The conclusions of a Bayesian analysis take probabilistic form, such as the result observed for VTE rates: "there is a 53 percent chance that the campaign reduced VTE rates by 5 percent or more in 2013." This flexibility of interpretation is depicted in Figure 4-25, which shows the estimates of HEN impact for VTE in 2013. Bayesian inference uses probabilistic language because the result of Bayesian analysis is a "posterior" probability distribution that describes the plausible range of values for the parameter of interest, in this case the impact of HEN alignment on VTE rates in 2013. Given this distribution, the probability that an impact exceeds a certain threshold corresponds to the area under the curve relative to that threshold, as Figure 4-25 shows. Based on this figure, the probability of at least a 5 percent reduction in VTEs is 53 percent, the area under the Bayesian posterior distribution to the left of the -5 percent change, corresponding to a reduction of at least 0.20 VTEs per 1,000 discharges. If estimated probabilities show more than 50 percent probability of a reduction in VTEs, it is more likely than not that the HEN component reduced VTEs. The closer the probabilities are to 100 percent, the more certain it is that an estimated impact of HEN alignment is likely to have occurred.

The probability of an event (e.g., a reduction in VTEs) and the probability of the complementary event (e.g., an increase in VTEs) must sum to 100 percent. When interpreting the probabilistic impacts of HEN alignment, it is thus important to consider the estimates' implications for the undesirable outcome: an increase in VTE rates. In Figure 4-25, the shaded area to the left of zero represents the probability of any reduction in VTEs, that is, an impact on VTEs falling below zero, which is depicted as a vertical line. The complement, the unshaded area, represents the probability that HEN alignment *increased* VTEs. Considering the complement of the desired impact is especially relevant when estimated probabilities of a given impact are close to 50 percent. For example, there is a 53 percent chance that the campaign reduced VTEs by 5 percent or more of the comparison group baseline in 2013, as shown in Figure 4-25; conversely, there is a 47 percent chance that the program was less effective (a decrease in VTEs between 0 and 5 percent) or that it increased VTEs. If the estimated probability is equal to 50 percent, an increase in VTE rates is as likely as a decrease.

This probabilistic framework extends to the interpretation of uncertainty intervals. The Bayesian uncertainty interval, shown immediately under the probability curve in Figure 4-25, defines the range within which the true parameter lies with 95 percent probability. The interpretation of the Bayesian uncertainty interval is as follows: "There is a 95 percent probability that the true impact of the HEN component of PfP falls between a 16.7 percent reduction and an 8.1 percent increase in the rate of VTEs." As in the frequentist framework, an uncertainty interval that spans zero suggests that zero is a plausible value for the parameter of interest. When

interpreting a given set of results, it is up to the policymakers to decide whether the results provide a sufficient level of confidence to make decisions about the program's future.

Figure 4-25—Interpreting the Bayesian Posterior Probability Distribution (for Impacts on VTE Rates in 2013)

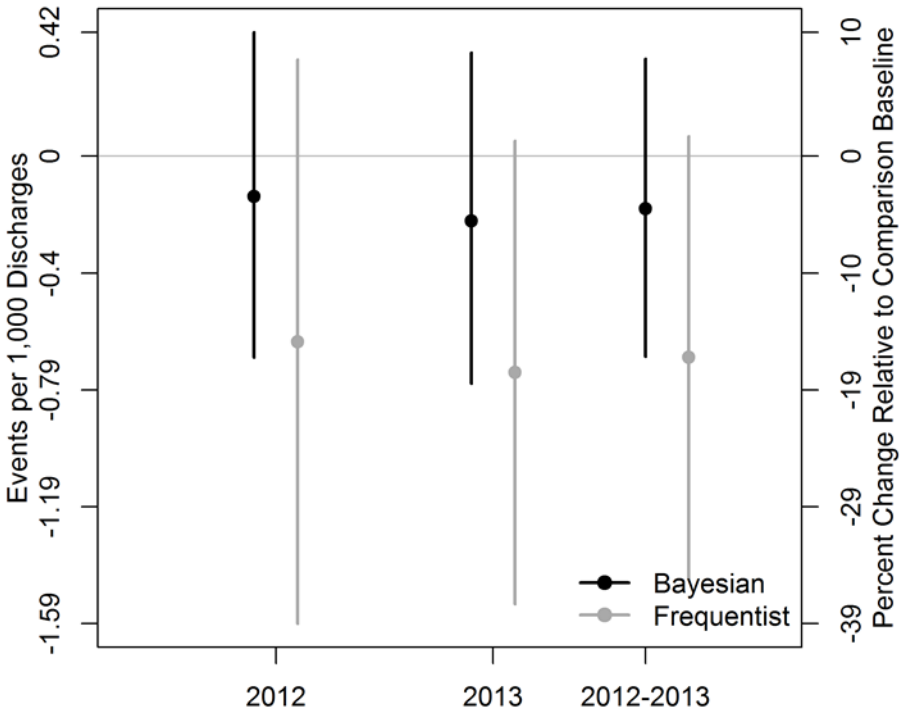


Source: Analysis of American Hospital Association (AHA) Annual Survey (fiscal year [FY] 2010) and Medicare inpatient claims data.
Note: This figure depicts the posterior distribution for the impact of HEN alignment on VTEs in 2013. Percent change is relative to the comparison baseline VTE rate (4.08 per 1,000), so a 5 percent reduction in VTEs corresponds to 0.20 VTEs averted per 1,000 discharges. Because the most recent quarter of data available spanned the first quarter of 2014, year 2013 includes all four quarters of 2013 and the first quarter of 2014.

Impact on VTEs for All HEN-Aligned Hospitals

Bayesian analysis results indicate a moderate probability of a slight improvement in VTE rates due to HEN alignment. Across both years of follow-up (2012–2013), HEN alignment was associated with a reduction of 0.18 VTEs per 1,000 discharges, corresponding to approximately 4.4 percent of the comparison baseline value of 4.08 VTEs per 1,000 discharges. The "comparison baseline" threshold values represent the average VTE rate among non-HEN-aligned hospitals in 2011 (Figure 4-26).

Figure 4-26—Estimated Effects and Uncertainty Intervals of HEN Alignment on VTEs



Source: Analysis of AHA Survey (FY 2010) and Medicare inpatient claims data.
Notes: Because the most recent data available spanned the first quarter of 2014, year 2013 includes all four quarters of 2013 and the first quarter of 2014. Dots indicate the point estimates of impact. The grey line represents traditional 95 percent confidence intervals, while the black lines represent Bayesian 95 percent credible intervals, where the endpoints are calculated as the 2.5 percentile and 97.5 percentile of the posterior probability distribution.

The Bayesian estimates indicate a smaller decrease in VTE rates than the frequentist estimates; however, for both sets of estimates the uncertainty interval around the impact estimate includes zero, so it is possible that the observed improvements in VTE rates are due to chance variation rather than HEN alignment. The bounds of the 95 percent Bayesian uncertainty interval are (-0.68, 0.33); there is a 95 percent chance that the true impact of HEN alignment on VTEs is between a reduction of 0.68 VTEs per 1,000 discharges and an increase of 0.33 VTEs per 1,000 discharges. As described above, this corresponds to -16.7 percent and 8.1 percent change in the comparison baseline value, respectively.

The frequentist results indicate that at the 5 percent level one cannot reject the hypothesis that there are no improvements in VTE rates due to HEN alignment, but that one can reject it at the 10 percent level. However, the Bayesian analysis leads to a more nuanced conclusion: it is highly likely that the HEN component is associated with small improvements in VTE rates (i.e., small reductions in VTE rates) in the post-intervention period, but that substantial improvements are unlikely. The first five rows of Table 4-2 present the probability that HEN alignment reduced VTEs at all, by 2 percent, 5 percent, 10 percent, or 25 percent of the comparison baseline value of 4.08 VTEs per 1,000 discharges in 2011. For example, there is a 76.3 percent chance that HEN alignment decreased VTE rates between 2011 and 2013, but only an 17.6 percent chance that HEN alignment decreased VTE rates by 10 percent or more of the comparison baseline, or 0.41 VTEs per 1,000 discharges, over the same period. Probabilities of small reductions in VTE rates (2 to 10 percent) are somewhat higher in 2013 than in 2012, but the probability of substantial reductions remain approximately zero.

Table 4-2—Probability of Impacts of HEN Alignment on VTE Rates, by Year

Change in VTE Rate Relative to Comparison Baseline (4.08 per 1,000)	Magnitude of Impact (VTEs per 1,000 Discharges)	Probability of Impact In		
		2012	2013	2012–2013
Any Decrease	< 0.00	0.691	0.789	0.763
Decrease of 2 Percent or More	≤ -0.08	0.574	0.699	0.657
Decrease of 5 Percent or More	≤ -0.20	0.404	0.527	0.462
Decrease of 10 Percent or More	≤ -0.41	0.171	0.245	0.176
Decrease of 25 Percent or More	≤ -1.02	0.001	0.002	0.000
Any Increase	> 0.00	0.309	0.211	0.237
Increase of 2 Percent or More	≥ 0.08	0.216	0.136	0.149
Increase of 5 Percent or More	≥ 0.20	0.113	0.063	0.061
Increase of 10 Percent or More	≥ 0.41	0.029	0.017	0.012
Increase of 25 Percent or More	≥ 1.02	0.000	0.000	0.000
Sample Sizes		**2012**	**2013**	**2012–2013**
Number of Treatment Hospitals	N/A	2,935	2,965	3,059
Number of Comparison Hospitals	N/A	742	733	808

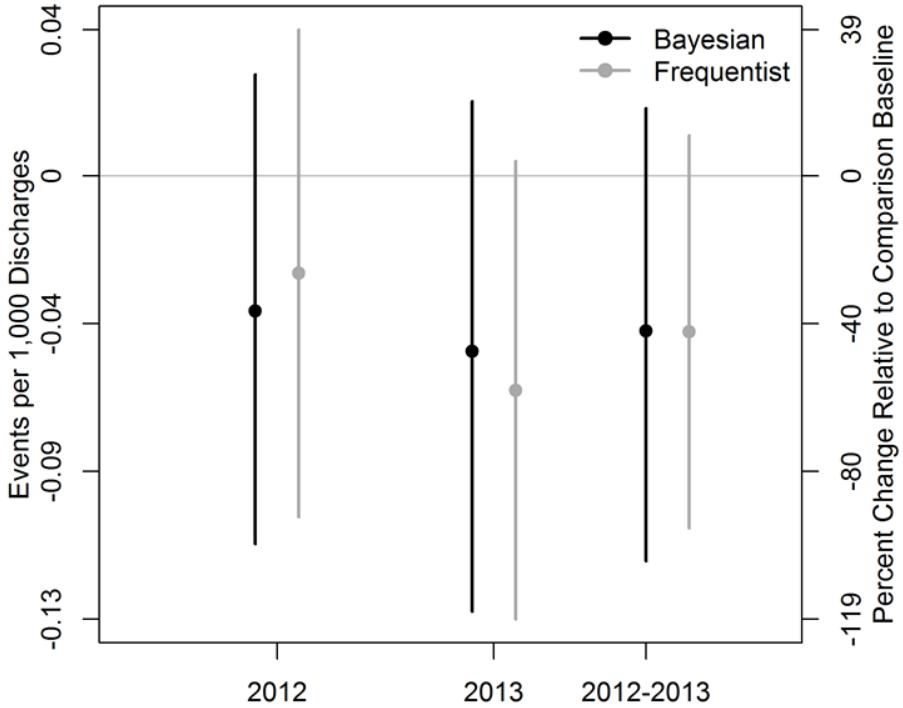
Source: Analysis of AHA Survey (FY 2010) and Medicare inpatient claims data.
Notes: Because the most recent data available spanned the first quarter of 2014, year 2013 includes all four quarters of 2013 and the first quarter of 2014 in both analyses.
Values reflect the probability that the impact estimate for the HEN component of PfP in the given year, calculated as the difference between that year and the 2011 baseline, is less than or equal to some threshold value. The "comparison baseline" threshold values represent the average VTE rate among non-HEN-aligned hospitals in 2011, or approximately 4.08 VTEs per 1,000 eligible discharges.

HEN alignment may also have increased the VTE rate among participating hospitals, so the last five rows of Table 4-2 quantify the complementary probabilities (any increase in VTEs or increases of 2 percent, 5 percent, 10 percent, or 25 percent). As with "any decrease," the probability of "any increase" in VTEs describes the chance that HEN alignment's impact is greater than zero. The probability that HEN alignment increased VTE rates by 5 percent or more of the comparison group baseline, that is, by 0.20 VTEs per 1,000 discharges, is 6 percent for the entire follow-up period. HEN alignment is not likely to have either reduced or increased VTE rates substantially.

Impact on Pressure Ulcers for All HEN-Aligned Hospitals

Bayesian analysis results indicate a moderate probability of a substantial impact on pressure ulcers due to HEN alignment, with larger reductions in 2013 than in 2012. Across both years of follow-up (2012–2013), HEN alignment was associated with an estimated reduction of 0.05 pressure ulcer events per 1,000 discharges, with an uncertainty interval of (-0.11, 0.02). Relative to the comparison value of 0.11 pressure ulcers per 1,000 discharges at baseline (2011), this impact represents a 41.9 percent decrease (Figure 4-27).

Figure 4-27—Estimated Effects and Uncertainty Levels of HEN Alignment on Pressure Ulcers



Source: Analysis of AHA Survey (FY 2010) and Medicare inpatient claims data.
Notes: Because the most recent data available spanned the first quarter of 2014, year 2013 includes all four quarters of 2013 and the first quarter of 2014.
Dots indicate the point estimates of impact. The grey lines represent traditional 95 percent confidence intervals, while the black lines represent Bayesian 95 percent uncertainty intervals, where the endpoints are calculated as the 2.5 percentile and 97.5 percentile of the posterior probability distribution.

The Bayesian uncertainty interval suggests that the true impact of HEN alignment falls somewhere between a complete reduction (down to zero) in pressure ulcers and an 18.2 percent increase. Table 4-3 reaffirms the notion that these impact estimates likely reflect a substantial impact of HEN alignment that increases slightly between 2012 and 2013.

Table 4-3—Probability of Impacts of HEN Alignment on Pressure Ulcers, by Year				
Change in Pressure Ulcer Rate Relative to Comparison Baseline (0.11 per 1,000)	**Magnitude of Impact (Pressure Ulcers per 1,000 Discharges)**	**Probability of Impact In**		
		2012	**2013**	**2012–2013**
Any Decrease	< 0.000	0.865	0.913	0.916
Decrease of 2 Percent or More	≤ -0.002	0.850	0.904	0.906
Decrease of 5 Percent or More	≤ -0.005	0.823	0.890	0.890
Decrease of 10 Percent or More	≤ -0.011	0.788	0.859	0.856
Decrease of 25 Percent or More	≤ -0.027	0.637	0.739	0.711
Any Increase	> 0.000	0.136	0.087	0.084
Increase of 2 Percent or More	≥ 0.002	0.120	0.077	0.076
Increase of 5 Percent or More	≥ 0.005	0.104	0.068	0.061
Increase of 10 Percent or More	≥ 0.011	0.081	0.049	0.043
Increase of 25 Percent or More	≥ 0.027	0.029	0.020	0.014
Sample Sizes		**2012**	**2013**	**2012–2013**
Number of Treatment Hospitals	N/A	3,177	3,163	3,246
Number of Comparison Hospitals	N/A	875	881	948

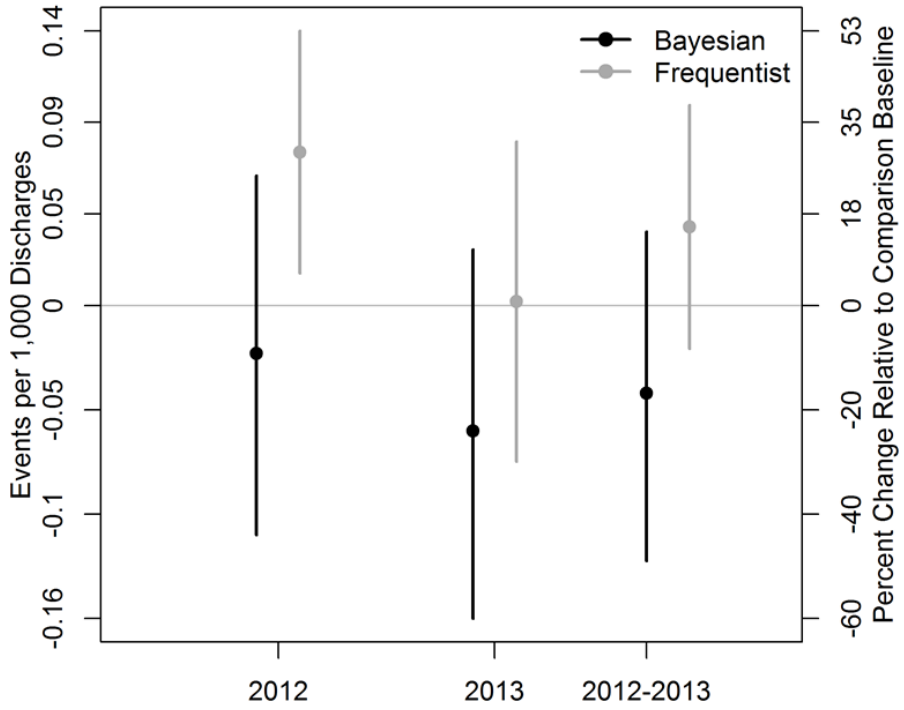
Source: Analysis of AHA Survey (FY 2010) and Medicare inpatient claims data.
Notes: Because the most recent data available spanned the first quarter of 2014, year 2013 includes all four quarters of 2013 and the first quarter of 2014 in both analyses.
Values reflect the probability that the impact estimate for the HEN component of PfP in the given year, calculated as the difference between that year and the 2011 baseline, is less than or equal to some threshold value. The "comparison baseline" threshold values represent the average pressure ulcer rate among non-HEN-aligned hospitals in 2011, or approximately 0.11 pressure ulcers per 1,000 eligible discharges.

The frequentist estimate for the full follow-up period is nearly identical to the Bayesian estimate. However, the frequentist results indicate that at the 10 percent level, one cannot reject the hypothesis that there are no improvements in pressure ulcer rates due to HEN alignment. The Bayesian results emphasize the high likelihood of an impact: the probability of at least a 10 percent decrease in pressure ulcers over the full post-intervention period, relative to the comparison baseline value of 0.11 pressure ulcers per 1,000 discharges, exceeds 85 percent (Table 4-2). Conversely, the probability of a 2 percent increase in pressure ulcers associated with HEN alignment is only 8 percent for the entire follow-up period, underscoring the strong probability that HEN alignment reduced pressure ulcers by a substantial margin.

Impact on CRBSI for All HEN-Aligned Hospitals

Bayesian analysis results indicate a moderate probability of small reductions in CRBSI rates due to HEN alignment. Across both years of follow-up (2012–2013), HEN alignment reduced CRBSI rates by an estimated 0.04 events per 1,000 eligible discharges, or 16.8 percent of the comparison baseline value of 0.26 events per 1,000 eligible discharges. The uncertainty interval of (-0.13, 0.04) events per 1,000 discharges places the impact between a decrease of 0.13 events per 1,000 discharges and an increase of 0.04 events per 1,000 discharges, corresponding to a decrease of 48.9 percent and an increase of 14.1 percent relative to the comparison baseline value (Figure 4-28).

Figure 4-28—Estimated Effects and Uncertainty Levels of HEN Alignment on CRBSI



Source: The Evaluation Contractor's Analysis of AHA Survey (FY 2010) and Medicare inpatient claims data.
Notes: Because the most recent data available spanned the first quarter of 2014, year 2013 includes all four quarters of 2013 and the first quarter of 2014.
Dots indicate the point estimates of impact. The grey lines represent traditional 95 percent confidence intervals, while the back lines represent Bayesian 95 percent credible intervals, where the endpoints are calculated as the 2.5 percentile and 97.5 percentile of the posterior probability distribution.

In the post-intervention period (2012–2013), there was an 82.8 percent chance that HEN alignment decreased CRBSI rates by 2 percent or more, but only a 66.3 percent chance that the program reduced adverse events by 10 percent or more, as Table 4-4 shows.

Table 4-4—Probability of Impacts of HEN Alignment on CRBSI Rates, by Year

Change in CRBSI Rate Relative to Comparison Baseline (0.26 per 1,000)	Magnitude of Impact (CRBSIs per 1,000 Discharges)	Probability of Impact In		
		2012	2013	2012–2013
Any Decrease	< 0.000	0.698	0.912	0.860
Decrease of 2 Percent or More	≤ -0.005	0.650	0.892	0.828
Decrease of 5 Percent or More	≤ -0.013	0.590	0.855	0.773
Decrease of 10 Percent or More	≤ -0.026	0.477	0.789	0.663
Decrease of 25 Percent or More	≤ -0.065	0.191	0.477	0.297
Any Increase	> 0.000	0.303	0.088	0.141
Increase of 2 Percent or More	≥ 0.005	0.261	0.073	0.117
Increase of 5 Percent or More	≥ 0.013	0.206	0.050	0.087
Increase of 10 Percent or More	≥ 0.026	0.131	0.027	0.044
Increase of 25 Percent or More	≥ 0.065	0.024	0.003	0.005
Sample Sizes		**2012**	**2013**	**2012–2013**
Number of Treatment Hospitals	N/A	3,147	3,199	3,269
Number of Comparison Hospitals	N/A	908	926	977

Source: Analysis of AHA Survey (FY 2010) and Medicare inpatient claims data.
Notes: Because the most recent data available spanned the first quarter of 2014, year 2013 includes all four quarters of 2013 and the first quarter of 2014 in both analyses.
Values reflect the probability that the impact estimate for the HEN component of PfP in the given year, calculated as the difference between that year and the 2011 baseline, is less than or equal to some threshold value. The "comparison baseline" threshold values represent the average CRBSI rate among non-HEN-aligned hospitals in 2011, or approximately 0.26 CRBSIs per 1,000 eligible discharges.

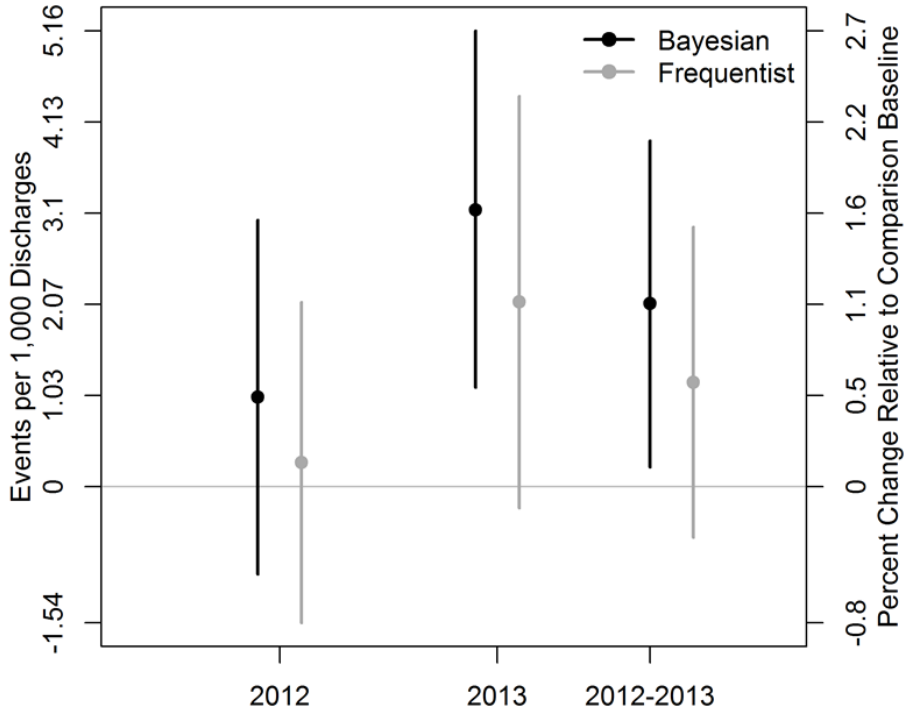
Across the entire follow-up period, there is an 86.0 percent chance that HEN alignment decreased CRBSI rates, implying a 14 percent chance that HEN alignment increased CRBSI rates. The probabilities of an increase in CRBSI rates in 2013 are substantially lower, with only a 5 percent chance of a 5 percent or greater increase in CRBSI rates.

Frequentist estimate in the full post-intervention period (2012–2013) indicates an increase in CRBSI rates; however, the effect is not statistically significant. Further, both the Bayesian and the frequentist uncertainty intervals span zero, indicating that it is possible that there was no effect of HEN alignment.

Impact on 30-Day Readmissions for All HEN-Aligned Hospitals

Bayesian analysis results provide no evidence for impact of HEN alignment. Across the follow-up period (2012-2013), HEN alignment is estimated to have increased readmission rates by 2.08 readmissions per 1,000 discharges, with an uncertainty interval of (0.22, 3.92). Compared to the comparison baseline of 188.9 readmissions per 1,000 discharges, the impact estimate represents a 1.1 percent increase, while the uncertainty interval suggests that the true impact lies between an increase of 0.12 percent and an increase of 2.1 percent (Figure 4-29). For practical purposes, this result can be summarized as "no evidence for HEN impact."

Figure 4-29—Estimated Effects and Uncertainty Intervals of HEN Alignment on 30-Day Readmissions



Source: The Evaluation Contractor's Analysis of AHA Survey (FY 2010) and Medicare inpatient claims data.
Notes: Dots indicate the point estimates of impact. The grey lines represent traditional 95 percent confidence intervals, while the black lines represent Bayesian 95 percent credible intervals, where the endpoints are calculated as the 2.5 percentile and 97.5 percentile of the posterior probability distribution.

The probability of decreased readmission rates dwindled from 15.5 percent in 2012 to less than 1 percent in 2013, or approximately 1.5 percent during the entire post-intervention period (Table 4-5). Although the probability of increased readmissions is very high at 99 percent or higher both overall and in 2013, the magnitude of the increase is quite small. There is only a 3.50 percent chance that HEN alignment increased readmissions by 2 percent or more across the post-intervention period. The frequentist estimate in the post-intervention period (2012–2013) is similar to the Bayesian estimate; however, it is not statistically significant at the 10 percent level, suggesting that one cannot reject the hypothesis of no impact.

Table 4-5—Probability of Impacts of HEN Alignment on Readmissions, by Year

Change in 30-Day Readmission Rate Relative to Comparison Baseline (188.9 per 1,000)	Magnitude of Impact (Readmissions per 1,000 Discharges)	Probability of Impact In		
		2012	2013	2012–2013
Any Decrease	< 0.00	0.155	0.001	0.015
Decrease of 2 Percent or More	≤ -3.78	0.000	0.000	0.000
Decrease of 5 Percent or More	≤ -9.45	0.000	0.000	0.000
Decrease of 10 Percent or More	≤ -18.89	0.000	0.000	0.000
Decrease of 25 Percent or More	≤ -47.23	0.000	0.000	0.000
Any increase	> 0.00	0.845	0.999	0.985
Increase of 2 Percent or More	≥ 3.78	0.004	0.263	0.035
Increase of 5 Percent or More	≥ 9.45	0.000	0.000	0.000
Increase of 10 Percent or More	≥ 18.89	0.000	0.000	0.000
Increase of 25 Percent or More	≥ 47.23	0.000	0.000	0.000
Sample Sizes		**2012**	**2013**	**2012–2013**
Number of Treatment Hospitals	N/A	3,462	3,441	3,463
Number of Comparison Hospitals	N/A	1,049	1,023	1,051

Source: Analysis of AHA Survey (FY 2010) and Medicare inpatient claims data.
Notes: Values reflect the probability that the impact estimate for the HEN component of PfP in the given year, calculated as the difference between that year and the 2011 baseline, is less than or equal to some threshold value. The "comparison baseline" threshold values represent the average 30-day readmissions rate among non-HEN-aligned hospitals in 2011, or approximately 188.9 readmissions per 1,000 discharges.

Taken together, these results for all outcomes suggest that hospitals that aligned with HENs did not consistently affect adverse events. Indeed, the differences between the HEN-aligned and non-HEN-aligned hospital outcomes are quite small. Of the four measures examined, the one substantial result is in pressure ulcers, where the outcome was reduced over time, and there is a greater than 85 percent probability that a reduction of over 10 percent occurred.

Impacts within Subgroups of Hospitals

Given the wide range of hospital types and implementation strategies that the PfP campaign encompassed, an analysis of impacts for all participating hospitals may gloss over crucial variation in impacts that result from this diversity. Identifying the types of hospitals and HENs in which HEN alignment produced improvements in outcomes is thus a key question for this evaluation. Estimating impacts in subgroups defined by hospital characteristics captures variation that may spring from hospitals' resources or organizational structure. Estimating impacts by HEN categories offers the ability to identify particularly promising or effective implementation strategies that might not be apparent in an overall analysis.

The Evaluation Contractor analyzed effects on VTE, pressure ulcers, CRBSI, and 30-day all-cause readmissions for 13 mutually exclusive hospital subgroups created based on hospital characteristics and for 10 mutually exclusive hospital subgroups based on characteristics of HENs with which they are aligned.

For subgroups of hospitals based on hospital characteristics, critical access hospitals (CAHs) were analyzed separately and the remaining 12 subgroups defined based on combinations of two hospital characteristics:

- Ownership (private, government, and nonprofit).
- Bed size (less than 100 beds, 100 to 199 beds, 200 to 399 beds, and 400 or more beds).

For example, the Evaluation Contractor examined impacts on CAHs, private hospitals with less than 100 beds, private hospitals with 100 to 199 beds, and so on for all 13 combinations.

The Evaluation Contractor also defined subgroups on the HENs' characteristics, referred to as HEN groups, hospitals were classified into system HENs and nine other subgroups of hospitals based on the following HEN characteristics:

- HEN type (hospital association and other).
- Intensity of one-on-one support provided to hospitals (high and low).
- Use of hospital collaboratives (yes or no).

System HENs (HENs managed by a hospital or healthcare system organization and designed to serve that system) were regarded as a separate category because of the uniqueness of system HENs in their ability to disseminate resources and provide direction on process changes across system hospitals. The non-system HENs needed to develop strategies to engage hospitals and support hospital changes through a variety of approaches. The Evaluation Contractor looked for implementation strategies that aligned with evidence-based principles of implementation and that distinguished HENs from one another. Two key implementation strategies emerged as both aligning with the literature and distinguishing HENs: (1) intensity of consultation and coaching to individual hospitals, and (2) use of collaboratives, or intensive, small groups of hospitals that committed to working together, piloting interventions, sharing best practices, and receiving coaching or technical assistance from the HEN. Appendix C provides additional detail about how HEN subgroups were constructed.

Table 4-6 and Table 4-7 map the subgroup abbreviations to their full descriptions.

Table 4-6—Hospital Subgroup Labels and Descriptions

Hospital Subgroup Label	Hospital Subgroup Description
CAH Hospital	Critical access hospital
Govt, < 100 Beds	Government (federal or non-federal)-owned hospital with 100 beds or fewer
Govt, 100–199 Beds	Government (federal or non-federal)-owned hospital with 100 to 199 beds
Govt, 200–399 Beds	Government (federal or non-federal)-owned hospital with 200 to 399 beds
Govt, >= 400 Beds	Government (federal or non-federal)-owned hospital with 400 or more beds
Nonprofit, < 100 Beds	Nonprofit hospital with 100 beds or fewer
Nonprofit, 100–199 Beds	Nonprofit hospital with 100 to 199 beds
Nonprofit, 200–399 Beds	Nonprofit hospital with 200 to 399 beds
Nonprofit, >= 400 Beds	Nonprofit hospital with 400 or more beds

Table 4-6—Hospital Subgroup Labels and Descriptions	
Hospital Subgroup Label	**Hospital Subgroup Description**
Private, < 100 Beds	Privately owned hospital with 100 beds or fewer
Private, 100–199 Beds	Privately owned hospital with 100 to 199 beds
Private, 200–399 Beds	Privately owned hospital with 200 to 399 beds
Private, >= 400 Beds	Privately owned hospital with 400 or more beds

Note: The hospital subgroups are mutually exclusive; that is, each hospital belongs to only one subgroup.

Table 4-7—HEN Group Labels and Descriptions	
HEN Group Label	**HEN Group Description**
Hosp Assn-High	Hospital association HEN that provided high-intensity support and did not use hospital collaboratives
Hosp Assn-High-Collab	Hospital association HEN that provided high-intensity support and used hospital collaboratives
Hosp Assn-Low	Hospital association HEN that provided low-intensity support and did not use hospital collaboratives
Hosp Assn-Low-Collab	Hospital association HEN that provided low-intensity support and used hospital collaboratives
Other-High	HENs not part of a hospital association or health system; HEN provided high-intensity support and did not use hospital collaboratives
Other-High-Collab	HENs not part of a hospital association or health system; HEN provided high-intensity support and used hospital collaboratives
Other-Low	HENs not part of a hospital association or health system; HEN provided low-intensity support and did not use hospital collaboratives
Other-Low-Collab[a]	HENs not part of a hospital association or health system; HEN provided low-intensity support and used hospital collaboratives
Other	Other HENs, not categorized by intensity level or participation in collaboratives.
System HEN	Health system HENs not categorized by intensity level or participation in collaboratives

Note: The HEN groups are mutually exclusive; that is, each HEN belongs to only one HEN group.
[a]Only one HEN fit this category.

These subgroups combine several characteristics of hospitals and HENs rather than considering separately a few discrete characteristics for conceptual and logistical reasons. The combined characteristics are expected to interact and potentially produce differences in the impact of an intervention. In addition, in order for the Bayesian model to draw information about impacts for one subgroup from the others, the subgroups need to contain more than five categories.

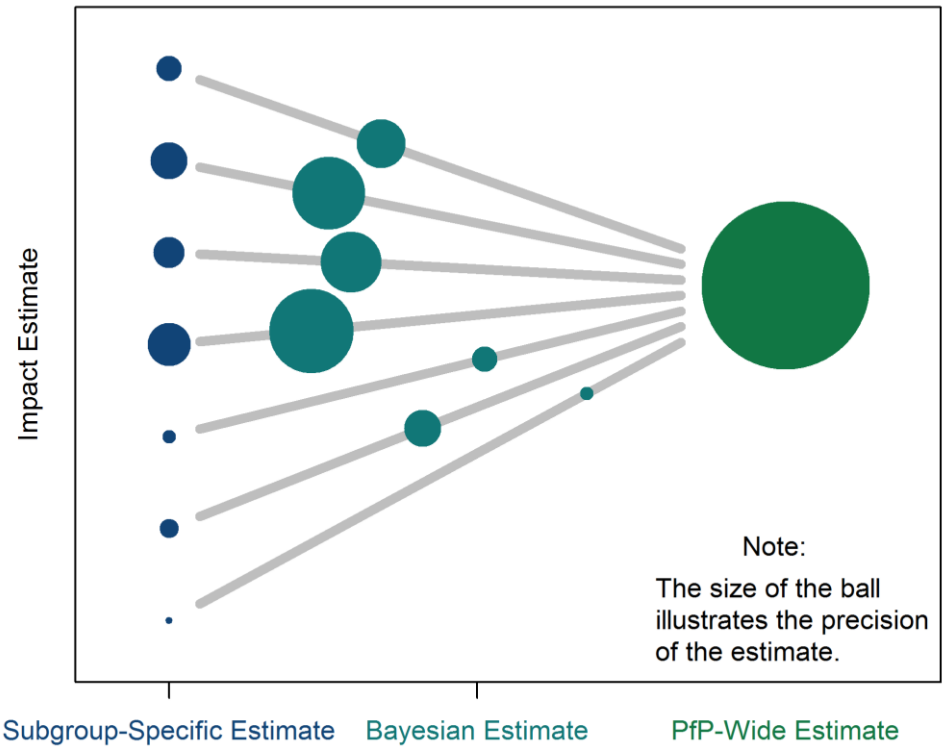
After a brief overview of Bayesian methods for subgroup analyses, results are presented for each outcome and subgroup, showing probabilities of increases and decreases in adverse event rates during the post-intervention period, with years 2012 and 2013 combined. Results for adverse events are shown first, followed by results for readmissions. Appendix E contains supplemental results for each outcome, including impact estimates, their standard errors, their uncertainty intervals, and probabilities of impacts of several magnitudes for each hospital and HEN subgroup.

In general, results for the post-intervention period (2012–2013) are consistent across hospital subgroups and HEN groups, with few differences in impacts across subgroups and few emerging trends. The relatively minor differences that occur are described below.

Bayesian Methods for Subgroup Analyses

To estimate the impact of HEN alignment on hospital and HEN subgroups, the Evaluation Contractor applied a Bayesian analytical model, which simultaneously reduces the likelihood of failing to detect an effect that is present and of falsely detecting an effect that is not present. As shown in Figure 4-30, Bayesian inference balances the extremes of no pooling, which examines each group individually, and complete pooling, which naïvely collapses all groups into a single unit. The Bayesian compromise of partial pooling estimates impacts as a weighted average of the overly noisy estimates for each group and the pooled estimates for all hospitals, balancing the information available from the data for each subgroup with the information available from the remaining subgroups.

Figure 4-30—Impact of Pooling Across Outcomes and Subgroups



Note: Illustrative sample data.

The Bayesian model allows the data to determine the appropriate degree of pooling. If the data suggest that differences among groups are negligible, that is, that the impact of the HEN component is relatively similar across hospital and HEN types, the model will pool more. By contrast, if the data suggest meaningful heterogeneity in the effects of the HEN component across groups, the model will pool less. In this way, the precision of the impact estimate for each subgroup of the dataset increases through drawing on information from the other subgroups, but only to the extent that the data deem them mutually informative (Gelman et al. 2013).

In contrast, the frequentist difference-in-differences regression approach analyzes each outcome, subgroup, and time period separately, which compounds issues with power, since estimating impacts for many subsets of data is likely to yield some statistically significant differences by chance alone. The risk of generating spuriously significant results is higher the more subgroups are analyzed with a goal of determining where, under what conditions, and for whom HEN alignment works. A unified Bayesian model reduces the likelihood of inferential errors due to multiple comparisons by more effectively separating "signal"—HEN alignment's true effect—from "noise"—random variability that masks underlying trends. This ability to "shrink" away noise obviates the need for multiple comparison adjustments, thereby facilitating the study of heterogeneous treatment effects across many subgroups.

Figure 4-31 through Figure 4-38 show probabilities of changes in outcomes for each hospital subgroup and HEN group analyzed. For hospital subgroups, the probability of impact among HEN-aligned hospitals was computed relative to non-HEN-aligned hospitals that belong to the same subgroup. For example, the probability of impact for HEN-aligned government hospitals was computed relative to non-HEN-aligned government hospitals. For HEN groups, the probability of impact among hospitals that belong to each HEN group is computed relative to all non-HEN-aligned hospitals, the reason being that non-HEN-aligned hospitals do not belong to a HEN and cannot be assigned a HEN type. For example, for hospitals that belong to a system HEN, probabilities of impacts were compared relative to all non-HEN-aligned hospitals.

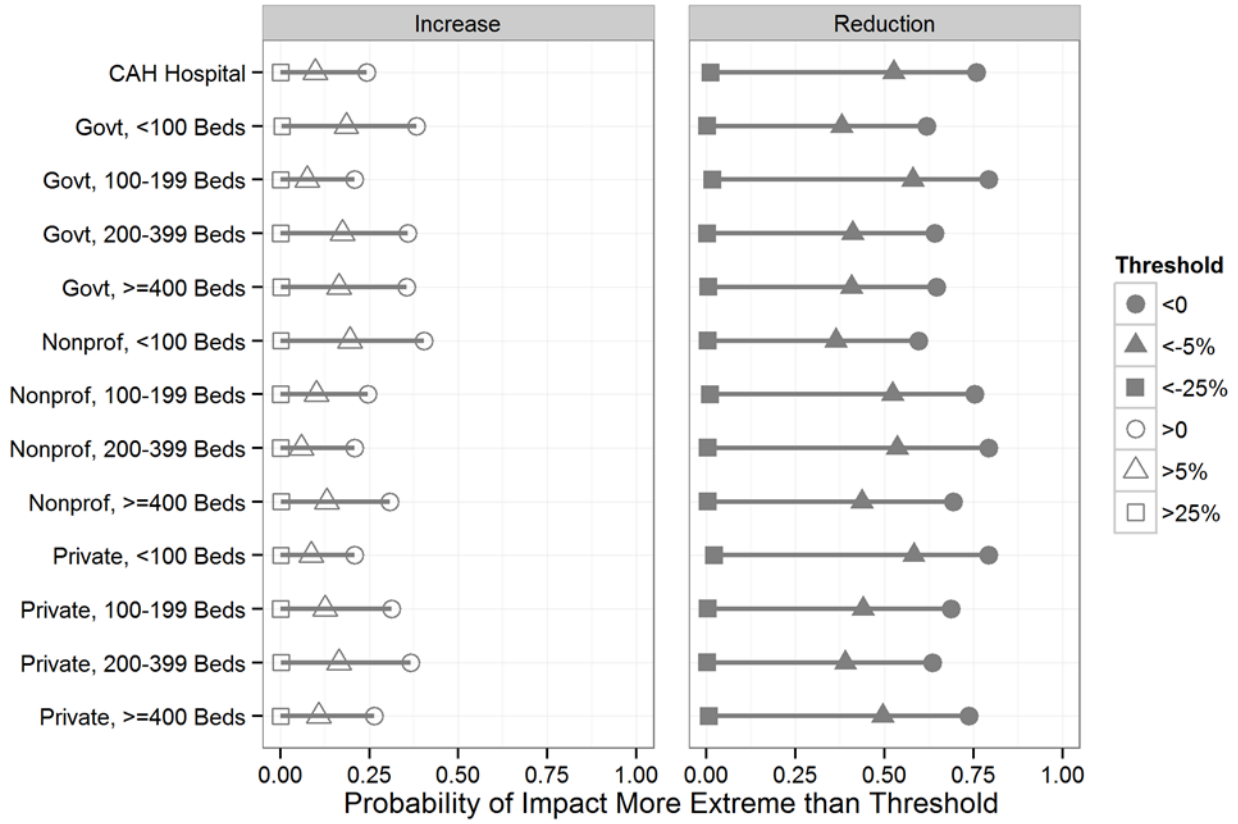
In each figure, the x-axis gives a probability and the labels on the y-axis denote hospital subgroups or HEN groups. The left-hand panel shows the probability that HEN alignment increased (worsened) the adverse event outcome at three thresholds: impact greater than zero, 5 percent (impact of 5 percent or more of the comparison baseline), and 25 percent (impact of 25 percent or more of the comparison baseline). The right-hand panel shows the probability that HEN alignment decreased the adverse event outcome at the same three thresholds, referring to impacts less than zero. In these figures, subgroup and HEN group names are abbreviated.

HEN Alignment's Impacts on VTEs Across Hospital Subgroups and HEN Groups

HEN alignment was associated with a moderate likelihood of improvement on VTEs across all hospital subgroups. The probability of a relatively small improvement associated with HEN alignment is considerably higher than the probability of a large improvement for all subgroups, while the probability of an increase in adverse events is low across hospital types. In summary:

- The probability of at least some decrease in VTEs associated with HEN alignment ranged from 60 to 79 percent.
- The probability of HEN alignment reducing VTEs by 25 percent or more was 2 percent or less for all hospital subgroups.
- No hospital subgroup had a 75 percent or higher likelihood of a 2 percent reduction or more associated with HEN alignment.
- No hospital subgroup has a statistically significant reduction of any size associated with HEN alignment.

Figure 4-31—HEN Alignment's Impacts on VTEs in the Post-Intervention Period (2012–2013), by Hospital Subgroup

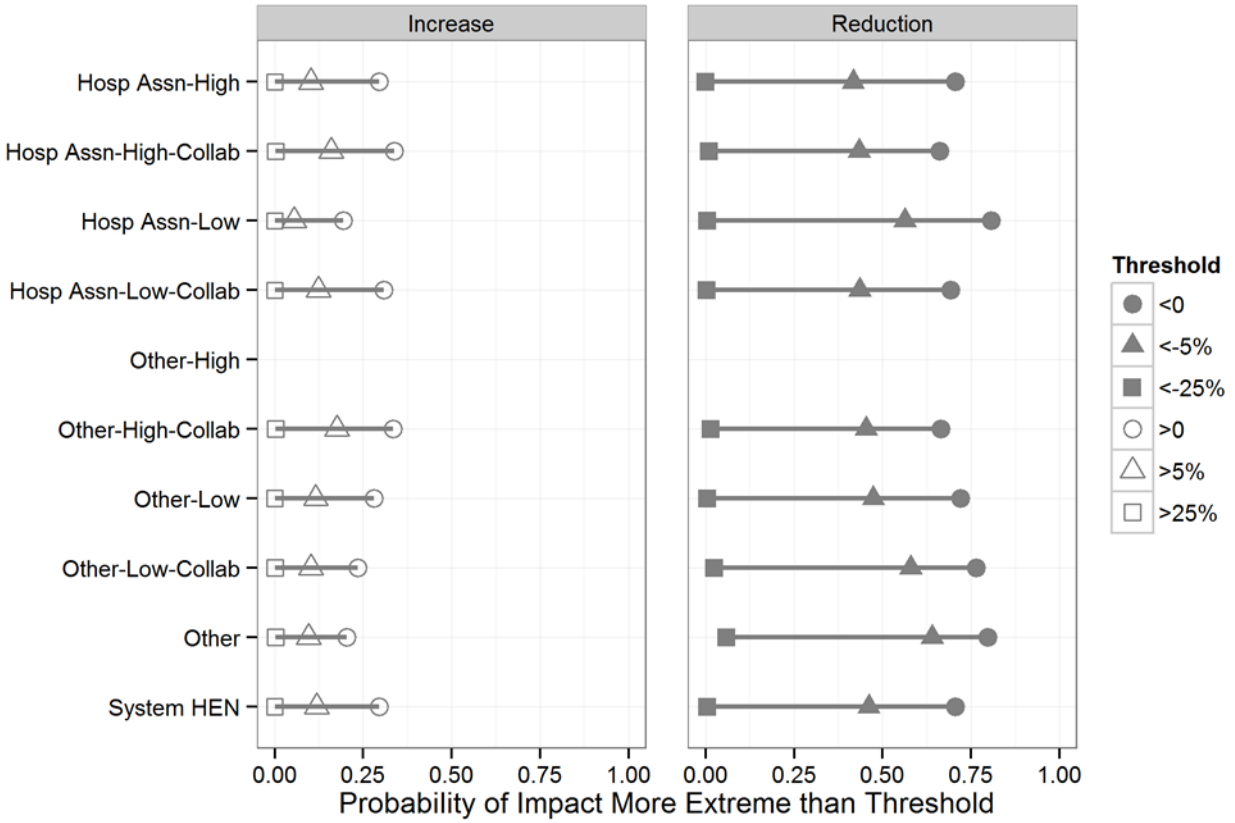


Source: Analysis of 2014 Survey on Prevention of Adverse Events and Readmissions, AHA Survey (FY 2010), and Medicare claims data.
Note: The comparison value for VTEs in 2011 is 4.08 events per 1,000 discharges. The subgroups are mutually exclusive; that is, each hospital belongs to only one subgroup.

Similar patterns appear for all HEN groups. Across all HEN groups, large reductions and large increases in VTEs associated with HEN alignment are improbable, but small reductions in VTEs are substantially more likely, without much variation across subgroups. In summary:

- The probability of any decrease in VTEs associated with HEN alignment ranged from 66 to 81 percent, while the probability of HEN alignment reducing VTEs by 25 percent or more ranged from less than 1 percent to 6 percent.
- No HEN subgroup had a 75 percent or higher likelihood of a 2 percent reduction or more associated with HEN alignment.
- No HEN subgroup has a statistically significant reduction of any size associated with HEN alignment.
- As Figure 4-32 shows, no hospitals in the "Other-High" HEN group (referring to high-intensity, non-collaborative HENs with an "other" organizational structure) provided sufficient data for analysis for this outcome.

Figure 4-32—HEN Alignment's Impacts on VTEs in the Post-Intervention Period (2012–2013), by HEN Group



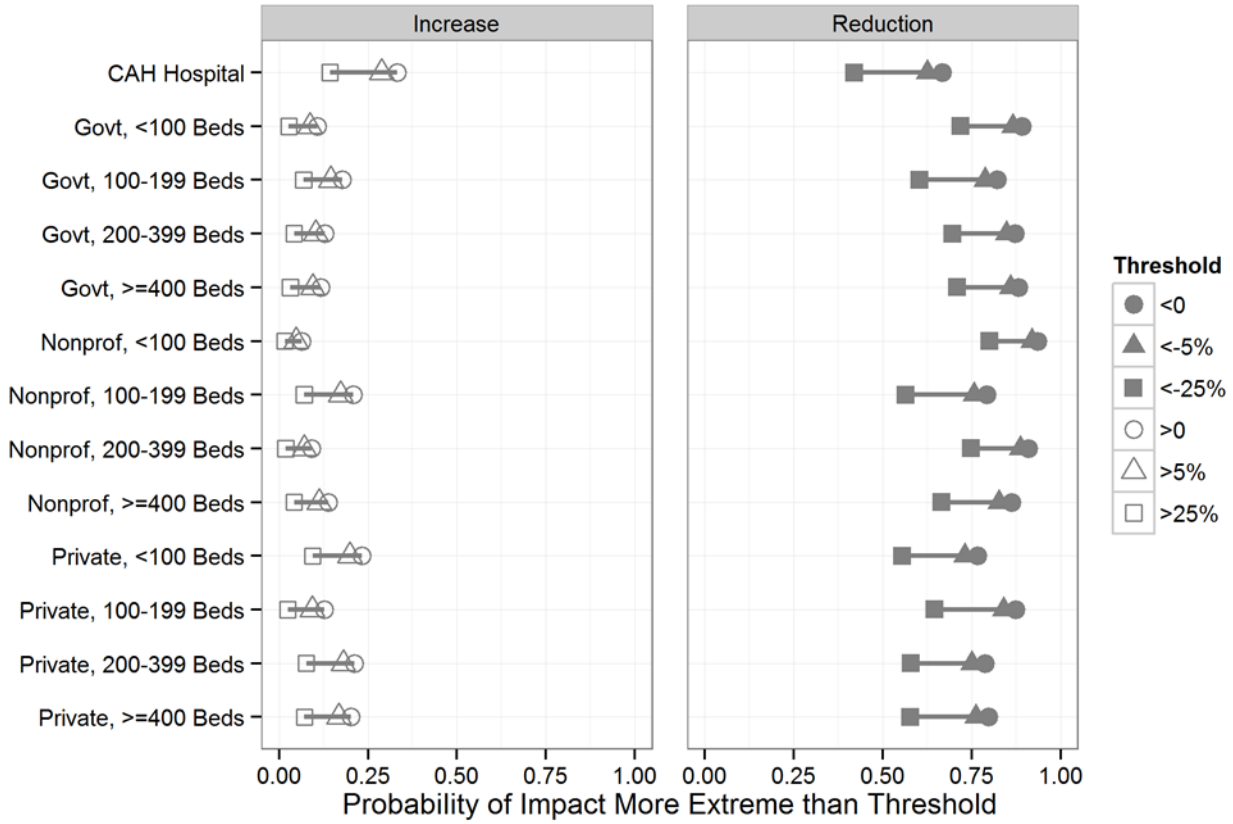
Source: Analysis of 2014 Survey on Prevention of Adverse Events and Readmissions, AHA Survey (FY 2010), and Medicare claims data.
Note: The comparison value for VTEs in 2011 is 4.08 events per 1,000 discharges. The HEN groups are mutually exclusive; that is, each HEN belongs to only one HEN group.

HEN Alignment's Impacts on Pressure Ulcers Across Hospital Subgroups and HEN Groups

The probability of HEN alignment reducing pressure ulcers was relatively high in all hospital subgroups and HEN groups in the post-intervention period, with few differences among subgroups. For hospital subgroups:

- The probability of any decrease in pressure ulcers associated with HEN alignment ranged from 67 to 94 percent, and the probability of reducing pressure ulcers by 25 percent or more ranged from 42 to 80 percent.
- All hospital subgroups except for CAHs and small private hospitals (less than 100 beds) had a 75 percent or higher likelihood of a 5 percent reduction or more associated with HEN alignment. CAHs stood out as having the lowest probability of a 5 percent or greater reduction in pressure ulcers at 63 percent, indicating that it was somewhat more likely than not that pressure ulcers were reduced by 5 percent or more associated with HEN alignment. In fact, in CAHs, the probability of an increase in pressure ulcers associated with HEN alignment was 33 percent.
- The likelihood of a 25 percent reduction or more associated with HEN alignment was 80 percent for small nonprofits (less than 100 beds).
- No hospital subgroup has a statistically significant reduction of any size associated with HEN alignment.

Figure 4-33—HEN Alignment's Impacts on Pressure Ulcers in the Post-Intervention Period (2012–2013), by Hospital Subgroup

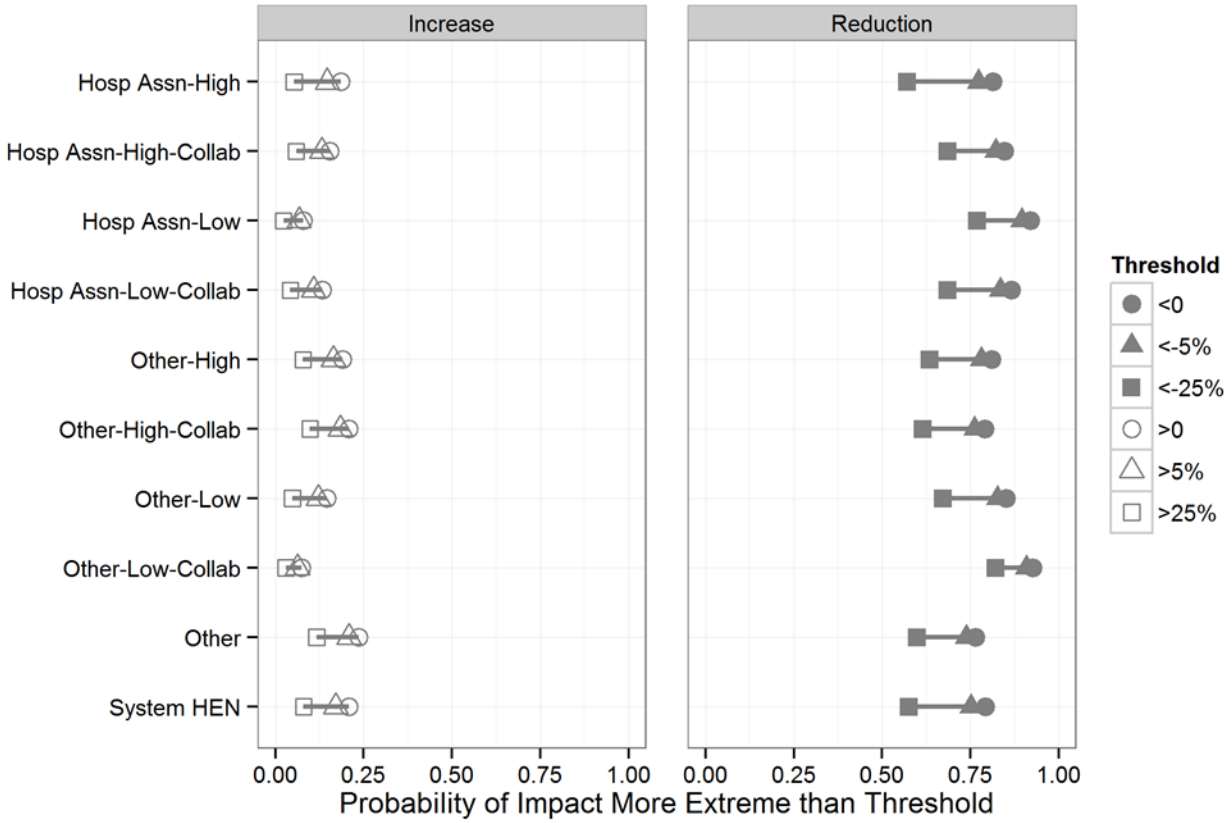


Source: Analysis of 2014 Survey on Prevention of Adverse Events and Readmissions, AHA Survey (FY 2010), and Medicare claims data.
Note: The comparison value for pressure ulcers in 2011 is 0.11 events per 1,000 discharges. The subgroups are mutually exclusive; that is, each hospital belongs to only one subgroup.

Among HEN groups, small differences are again seen across subgroups:

- The probability of any decrease in pressure ulcers associated with HEN alignment ranged from 76 to 93 percent, and the probability of HEN alignment reducing pressure ulcers by 25 percent or more ranged from 57 to 82 percent.
- The likelihood of a 25 percent reduction or more associated with HEN alignment was 75 percent or higher for the following subgroups: low-intensity collaborative "other" type HENs ("other" means those that were not system-owned or hospital association-based HENs, as described in Table 4-7) and low-intensity non-collaborative hospital association HENs.[4-35]
- No HEN subgroup had a statistically significant reduction of any size associated with HEN alignment.

Figure 4-34—HEN Alignment's Impacts on Pressure Ulcers in the Post-Intervention Period (2012–2013), by HEN Group



Source: Analysis of 2014 Survey on Prevention of Adverse Events and Readmissions, AHA Survey (FY 2010), and Medicare claims data.
Note: The comparison value for pressure ulcers in 2011 is 0.11 events per 1,000 discharges. The HEN groups are mutually exclusive; that is, each HEN belongs to only one HEN group.

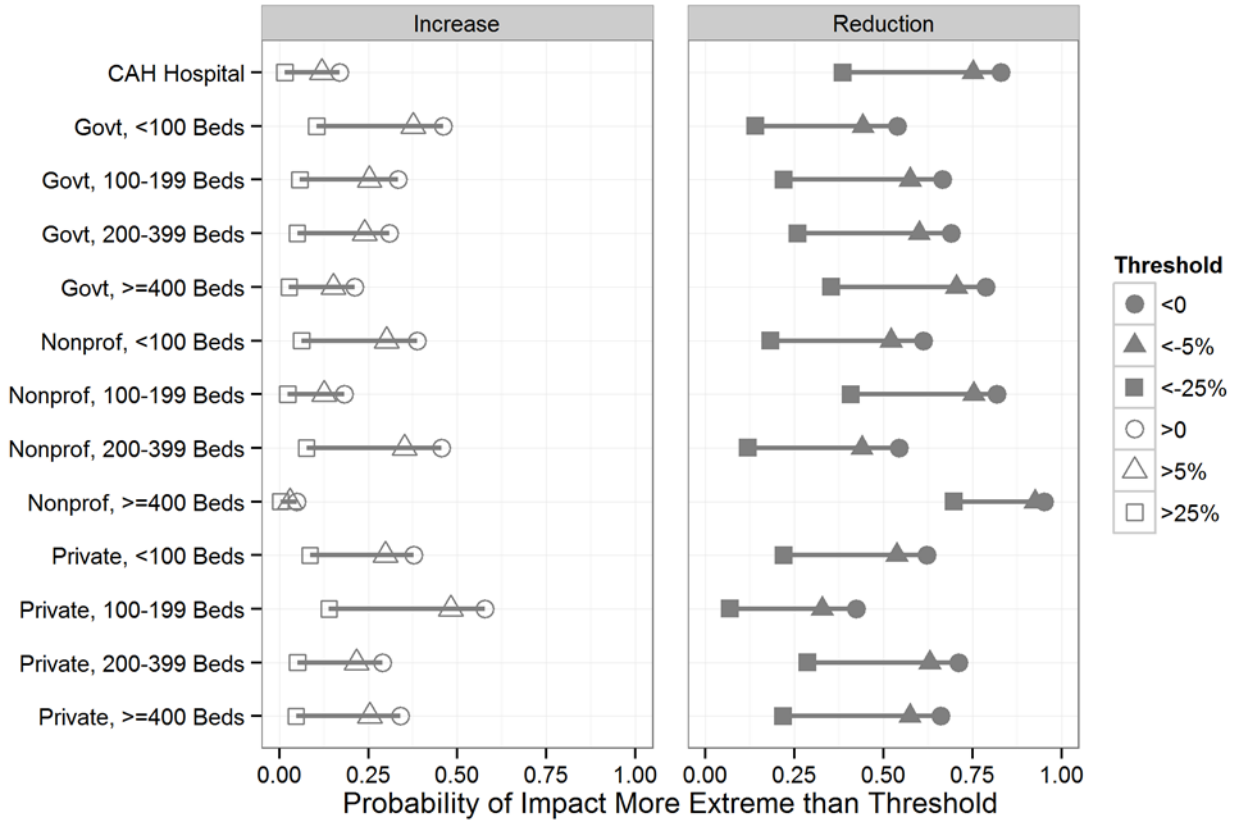
4-35 Only one HEN fit this category.

HEN Alignment's Impacts on CRBSI Across Hospital Subgroups and HEN Groups

Unlike the general pattern for most subgroups and outcomes analyzed, the probability of HEN impact on CRBSI rates varied widely by hospital subgroup, but without a unifying theme:

- The probability of any decrease in CRBSI rates associated with HEN alignment ranged from 42 to 95 percent, and the probability of HEN alignment reducing CRBSI rates by 25 percent or more ranged from 7 percent in moderately sized private hospitals (100 to 199 beds) to 70 percent in large nonprofit hospitals (400 or more beds). For moderately sized private hospitals (100 to 199 beds), there was a 58 percent chance of an *increase* in CRBSI associated with HEN alignment, indicating that it is slightly more likely than not that the increase occurred. However, larger than 50 percent probabilities were estimated only for 2 percent or larger increases.
- CAHs, large nonprofit hospitals (400 or more beds), and nonprofit hospitals with 100 to 199 beds had a 75 percent or larger probability of a 5 percent or greater reduction in CRBSI rates associated with HEN alignment.
- No subgroup of hospitals had a 75 percent or larger probability of a 25 percent reduction or more associated with HEN alignment.
- No hospital subgroup had a statistically significant reduction of any size associated with HEN alignment.

Figure 4-35—HEN Alignment's Impacts on CRBSI in the Post-Intervention Period (2012–2013), by Hospital Subgroup

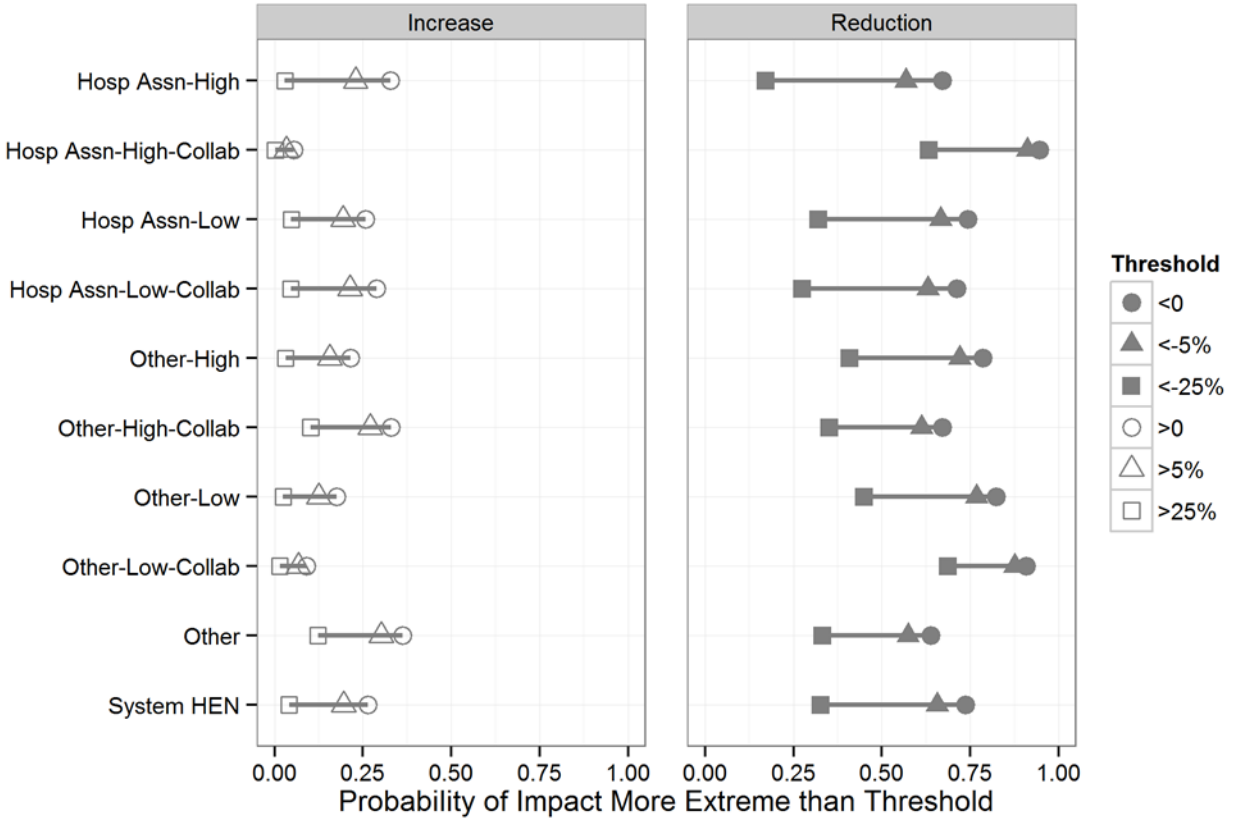


Source: Analysis of 2014 Survey on Prevention of Adverse Events and Readmissions, AHA Survey (FY 2010), and Medicare claims data.
Note: The comparison baseline value for CRBSI is 0.26 events per 1,000 discharges. The subgroups are mutually exclusive; that is, each hospital belongs to only one subgroup.

Across HEN groups, there was also moderate variation in impacts of HEN alignment across groups, again without a clear pattern:

- The probability of any decrease in CRBSI rates associated with HEN alignment ranged from 64 to 95 percent, and the probability HEN alignment of reducing CRBSI rates by 25 percent or more ranged from 17 to 69 percent.
- High-intensity collaborative hospital association HENs and low-intensity collaborative HENs had a 75 percent or larger probability of a 10 percent or greater reduction in CRBSI rates associated with HEN alignment. (Appendix E shows probabilities of a 10 percent or greater reduction.)
- No HEN groups had a 75 percent or larger probability of a 25 percent reduction or more associated with HEN alignment.
- No HEN subgroup had a statistically significant reduction of any size associated with HEN alignment.

Figure 4-36—HEN Alignment's Impacts on CRBSI in the Post-Intervention Period (2012–2013), by HEN Group



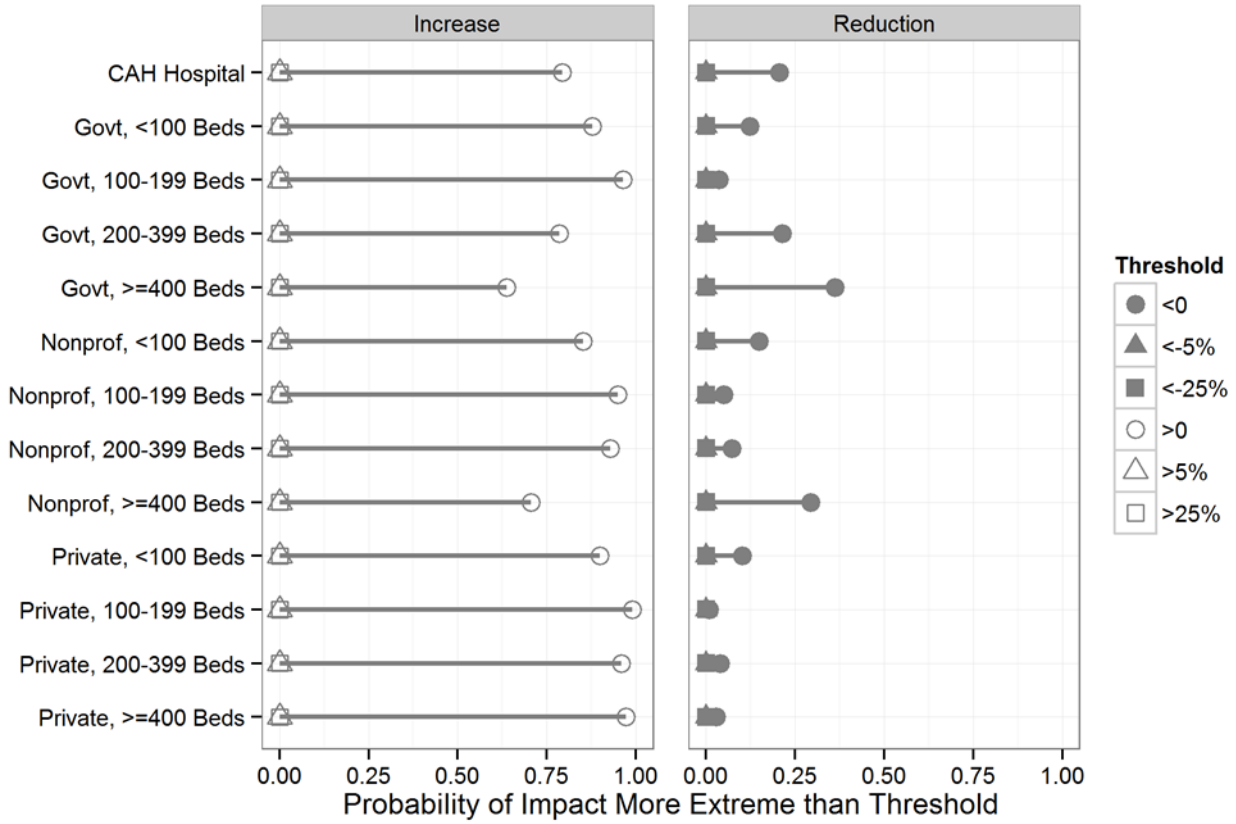
Source: Analysis of 2014 Survey on Prevention of Adverse Events and Readmissions, AHA Survey (FY 2010), and Medicare claims data.
Note: The comparison value for CRBSI in 2011 is 0.26 events per 1,000 discharges. The HEN groups are mutually exclusive; that is, each HEN belongs to only one HEN group.

HEN Alignment's Impacts on Readmissions Across Hospital Subgroups and HEN Groups

HEN alignment produced small but highly likely increases in 30-day readmissions in all hospital subgroups and HEN groups, with similar probabilities of an increase across subgroups:

- The probability that HEN alignment reduced readmissions by 5 percent or more of the comparison baseline was zero for all hospital subgroups in the post-intervention period.
- The probability of any increase in readmissions associated with HEN alignment was high—it was greater than 75 percent for all subgroups except for large government and nonprofit hospitals (more than 400 beds) (which were both above 60 percent).
- Although increases in readmissions are likely, these increases are not likely to be large; the probability of increases of 5 percent or more associated with HEN alignment was zero in all subgroups.
- In Private hospitals with 100-199 beds, there was a statistically significant (99.1 percent) probability that readmissions dropped less in HEN-aligned hospitals than in the comparison group.

Figure 4-37—HEN Alignment's Impacts on Readmissions in the Post-Intervention Period (2012–2013), by Hospital Subgroup

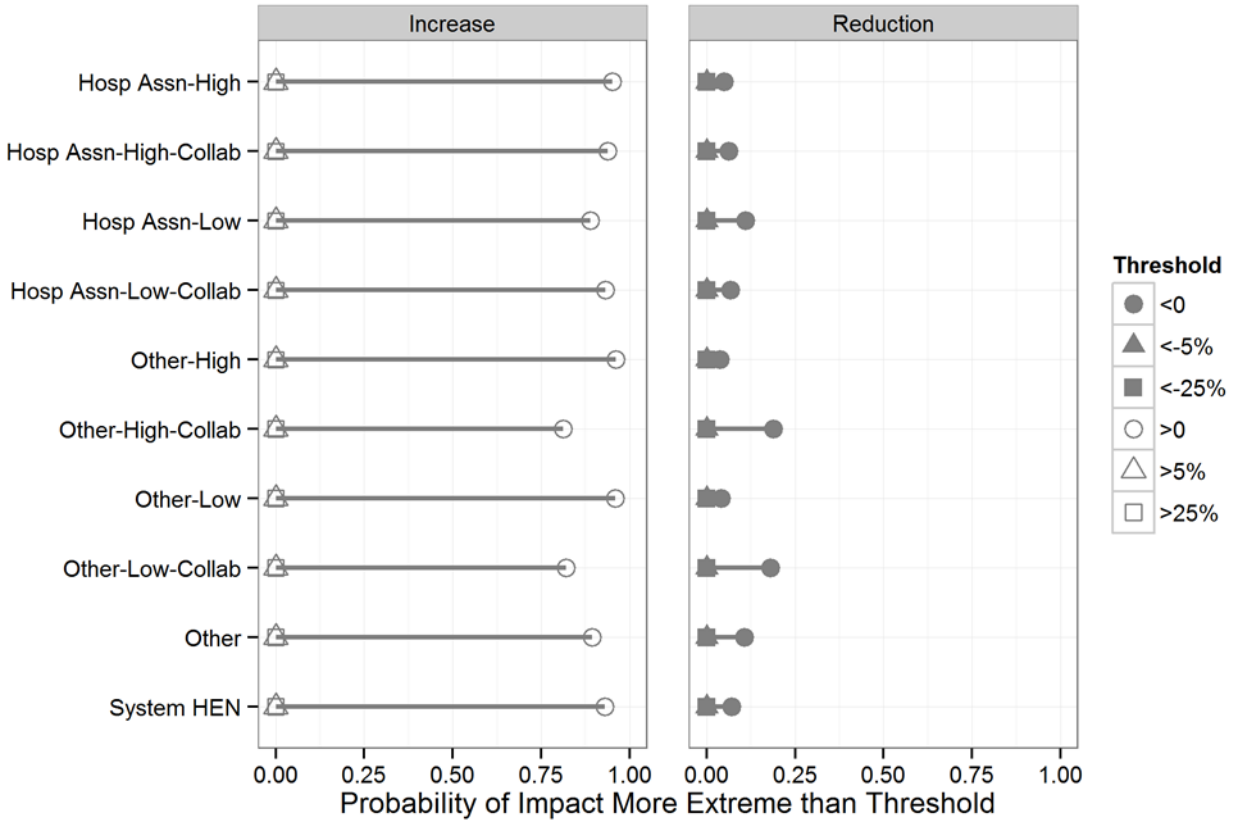


Source: Analysis of 2014 Survey on Prevention of Adverse Events and Readmissions, AHA Survey (FY 2010), and Medicare claims data.
Note: The comparison value for readmissions in 2011 is 188.9 readmissions per 1,000 discharges. The subgroups are mutually exclusive; that is, each hospital belongs to only one subgroup.

Similar results were obtained among HEN groups, with little variation across subgroups:

- The probability of HEN alignment reducing readmissions by 5 percent or more was zero in all HEN groups.
- There is a high probability of some increase in readmissions in all HEN groups associated with HEN alignment, but zero probability of increases in readmissions rates of 5 percent or more.
- No HEN subgroup had a statistically significant reduction of any size associated with HEN alignment.

Figure 4-38—HEN Alignment's Impacts on Readmissions in the Post-Intervention Period (2012–2013), by HEN Group



Source: Analysis of 2014 Survey on Prevention of Adverse Events and Readmissions, AHA Survey (FY 2010), and Medicare claims data.
Note: The comparison baseline value for readmissions is 188.9 readmissions per 1,000 discharges. The HEN groups are mutually exclusive; that is, each HEN belongs to only one HEN group.

Does Changing the Definition of the HEN-Aligned and Comparison Group Improve Our Ability to Identify Effects?

In addition to the direct effect the HENs may have had on HEN-aligned hospitals, it is possible that the national campaign, which enjoyed substantial visibility, altered the awareness of the importance of patient safety and inspired non-HEN-aligned hospitals to participate in PfP activities to reduce harms. As described in Chapter 1, reaching *all* hospitals was a goal of the campaign. As of summer 2013, 17 of the 26 HENs made their activities or resources available in some way to non-HEN-aligned hospitals.[4-36] Common ways included opening webinars and training to any hospital, and maintaining an online space in which their patient safety improvement resources are available to a broad audience.

In this analysis, the Evaluation Contractor re-examined the impact of the PfP campaign on readmissions and adverse events after changing the definition of HEN-aligned hospitals to include any hospitals (that is, any in the survey sample, for whom the Evaluation Contractor had data) that reported working with a HEN or state hospital association (SHA) on patient safety improvement, regardless of whether they were aligned with a HEN. Outcomes for that re-defined group were then compared to a re-defined comparison group which excluded hospitals that said they had been influenced by PfP as a national effort.

More specifically, in response to the 2014 Survey on Prevention of Adverse Events and Readmissions, hospitals self-reported whether or not they worked with a HEN or SHA to improve patient safety and reduce readmissions during any or all of 2012–2013. Hospitals that reported that they had not worked with a HEN were asked whether the national PfP campaign influenced their work on patient safety, and approximately 18 percent of these non-HEN hospitals responded that they had been influenced by the PfP campaign. After removing these hospitals from the comparison group, changes in outcomes among the re-defined HEN-aligned group were compared to non-HEN-aligned survey respondents who reported not being influenced by PfP using difference-in-differences models. Propensity-score reweighting was used to better assure comparability between the two groups on factors other than their history of work with a HEN.

Out of four outcomes, the analysis found only one significant difference-in-differences impact estimate. That is, the Evaluation Contractor found a statistically significant difference between the change in the VTE rate among the re-defined HEN group relative to the re-defined comparison group, and the size of the impact was relatively large (27.60 percent of the baseline VTE rate of the HEN-aligned group). The change in the estimated impact from the frequentist or Bayesian analyses discussed above stemmed from both the addition of hospitals to the HEN-aligned group who self-reported working with a HEN but who were not aligned as of June 2012, as well as exclusion from the non-HEN-aligned group of hospitals who indicated PfP had influenced them.

However, the impact estimates were not statistically significant for readmissions, CRBSI, or pressure ulcers (Table 4-8).[4-37] There is nothing in the qualitative assessment that suggests why there would be a difference in the effect for VTE but no other areas of harm. Taken together, these analyses suggest that HENs and PfP more broadly may have influenced hospitals beyond the HEN-aligned group, but to a limited degree.

[4-36] Based on HEN interviews conducted in summer 2013, which asked "to what extent were your Partnership activities also made available to non-HEN-aligned hospitals?"

[4-37] The result for readmissions is consistent with the frequentist HEN impact estimate shown in Figure 4-29, which uses similar methodology.

Table 4-8—Estimated Impacts of PfP on Medicare Patients' Adverse Event Rates after Removing Spillover

Outcome	Unadjusted Estimates							Regression-Adjusted Estimates
	HEN-Aligned Hospitals			Non-HEN-Aligned Hospitals			Difference-in-Differences (SE)	Difference-in-Differences Estimate (SE)
	2011 Baseline	2013 Follow-up	Difference	2011 Baseline	2013 Follow-up	Difference		
Readmissions[a] N=1,000 hospitals	17.62	16.11	-1.51	16.38	14.93	-1.45	-0.07 (0.4)	0.30 (0.64)
CRBSI (PSI- 07)[b] N=884 hospitals, 4,054,231 discharges	0.48	0.26	-0.22	0.45	0.19	-0.26	0.04 (0.09)	-0.04 (0.09)
Pressure Ulcers (PSI- 03)[b] N=955 hospitals, 2,094,758 discharges	0.64	0.54	-0.11	0.68	0.45	-0.22	0.12 (0.26)	0.03 (0.22)
VTE (PSI 12)[b] N=880 hospitals, 1,735,528 discharges	6.45	5.91	-0.54	4.66	5.90	1.25	-1.78* (0.85)	-1.86* (0.78)

Source: Evaluation Contractor's analysis of Medicare inpatient claims data.

Note: The difference-in-difference impact estimates shown in this table represent differences between the changes in outcomes of the HEN-aligned hospitals between baseline and follow-up relative to changes in the comparison hospitals over the same period.

Each row corresponds to a separate difference-in-differences regression using propensity score-based weights. The first six columns present the components of the difference-in-differences estimates using unadjusted adverse event and readmission rates. The seventh column shows the difference-in-differences estimate without regression adjustment. The final column presents the main impact estimate, adjusting for patient characteristics, patient risk factors, and hospital fixed effects. The impacts were estimated by using linear probability regression models. Robust standard errors (SEs), clustered by hospital, are reported in parentheses.

*Difference-in-differences treatment-comparison impact estimate significantly different from 0 at the 0.05 level, two-tailed test.

**Difference-in-differences treatment-comparison impact estimate significantly different from 0 at the 0.01 level, two-tailed test.

[a]30-day all-cause readmissions per 100 discharges

[b]Reported as the number of adverse events per 1,000 Medicare beneficiary discharges at risk

Impact Estimates Using Medicare Patient Safety Monitoring System (MPSMS) Data

Introduction

The MPSMS data served as another important data source for analysis of the key research questions. The advantage of this dataset is that it includes a large national sample of medical charts subjected to review by professional abstractors.[4-38] Adverse events are much more likely to be detected through abstraction of medical charts than through claims data. Because of this better detection, the rate of adverse events in medical chart review data is much higher than in claims data, and might provide a more accurate measure for purposes of distinguishing HEN impacts on extremely rare events.

However, the MPSMS data are drawn from a sample of hospital charts, rather than the complete national Medicare FFS claims database (which includes not a sample of patients but rather Medicare FFS patients across all United States [U.S.] hospitals). Thus, the statistical power to detect the effects of a national initiative to improve patient safety is more limited when using the MPSMS data than when using claims data, meaning the analysis is less likely to be able to detect an effect if it truly exists.

This section presents results from the Evaluation Contractor's analysis of 2009–2013 data from the MPSMS, a nationwide federal project for surveillance of hospital-associated adverse events sponsored by the AHRQ. MPSMS data for 2014 were not available in time to be included in this report. The abstraction process identifies adverse events in four categories: ADE (5 measures), post-procedural adverse events (7 measures), general adverse events (2 measures: falls and pressure ulcers), and hospital-acquired infections (HAIs) (5 measures)—for a total of 19 measures. A full list of the measures is in Appendix E.

In order to improve the likelihood of finding evidence regarding the impact of HEN alignment on adverse events, the Evaluation Contractor grouped the data into six composite measures: the total *number* of adverse events per 1,000 discharges, and five patient-level composite binary "any" variables—one for each of the four categories above (for any adverse event in that category), plus a fifth "overall any" outcome for any of the 19 adverse events at all. These six composite measures were: (1) the count of the number of adverse events per 1,000 discharges; (2) any ADE; (3) any post-procedural adverse event; (4) any "general" adverse event (fall or pressure ulcer); (5) any HAI; and (6) any adverse event at all. This analysis focuses on these six composite measures because of their greater incidence, and hence their relatively greater statistical power to detect impacts (compared to analyses of any single adverse event).

4-38 Health Services Advisory Group and Mathematica Policy Research. Project Evaluation Activity in Support of Partnership for Patients: Evaluation Annual Report. Submitted to the Centers for Medicare & Medicaid Services. Phoenix: Health Services Advisory Group and Washington, DC: Mathematica Policy Research, September 2014.
Clarkwest A, Kranker K, Witmer S, et al. Project Evaluation Activity in Support of Partnership for Patients: Task 2 Evaluation Plan. Report submitted to the Centers for Medicare and Medicaid Services. Phoenix: Health Services Advisory Group, and Washington, NJ: Mathematica Policy Research, April 2012.
Classen DC, Resar R, Griffin F, et al. Global trigger tool shows that adverse events in hospitals may be ten times greater than previously measured. *Health Affairs*. 2011; 30(4):581–589.
Office of the Inspector General, U.S. Department of Health and Human Services. Adverse events in hospitals: national incidence among Medicare beneficiaries. 2010b. Available at http://oig.hhs.gov/oei/reports/oei-06-09-00090.pdf. Accessed on: August 13, 2015.

Impact Analyses with MPSMS Data[4-39]

The regression-adjusted difference-in-differences impact analyses for each of the 19 adverse events in MPSMS, and for six composite variables (the one count variable and the five composite "any" variables) found no evidence of any impacts of HEN alignment. Table 4-9 presents the impact estimates for the six composite measures (impact estimates results for all 19 adverse event outcomes are in Appendix E). The estimates of impact were small and not statistically significant. Two point estimates were favorable (less than zero, indicating adverse events decreased more in the HEN-aligned hospitals than adverse events decreased in nonaligned hospitals) and four point estimates were unfavorable (greater than zero). In all, no measurable effects of HEN-alignment were detected using the MPSMS-based composite measures. However, given the relatively small sample sizes, the statistical power of these analyses is limited, as shown in Table 4-9. Thus, the null findings may be either because HEN alignment had no effects on these outcomes or because the statistical power to detect small (that is, 5 percent) effects was poor. The analysis for the main composite measure did have the statistical power to identify a large effect of 20 percent or more, and probably would have identified an effect of 10 percent or more (Table 4-9, first row, last two columns). Appendix E presents detailed results for all of the MPSMS outcomes that were analyzed.[4-40]

Table 4-9—Difference-in-Differences Impact Analyses For Composite "Any" Adverse Event Outcomes

Measure	Unadjusted							Regression-Adjusted Difference-in-Differences	Number of Observations	Power to Detect a 5 Percent Effect[a]	Power to Detect a 10 Percent Effect[a]	Power to Detect a 20 Percent Effect[a]
	Non-HEN-Aligned Hospitals			HEN-Aligned Hospitals			Difference-in-Differences					
	Pre 2009–2011	Post 2012–2013	Difference	Pre 2009–2011	Post 2012–2013	Difference						
Any adverse event[b]	13.74	11.92	-1.81 (0.58)**	14.79	13.47	-1.33 (0.27)**	0.49 (0.63)	**0.44 (0.53)**	125,004	25.1	73.2	99.9
Count of the number of adverse events[c] (per 1,000 cases)	183.41	157.15	-26.26 (9.29)**	199.57	177.71	-21.87 (4.14)**	4.39 (10.17)	**-4.02 (8.92)**	125,004	17.5	53.8	98.4
Any ADE[d]	6.85	6.09	-0.75 (0.41)	7.45	6.50	-0.95 (0.19)**	-0.20 (0.45)	**-0.20 (0.42)**	117,234	12.5	36.8	90.1

4-39 The data used in this analyses are the full sample of medical chart review data from MPSMS. Therefore, they are different from the AHRQ National Scorecard HAC rates presented in Chapter 2, which were adjusted to be nationally representative rather than focused on patients with four high-risk conditions, and which included measurement of several HACs from other data sources.

4-40 The impact analyses for the 19 separate measures produced one statistically significant result indicating a higher rate for Mechanical Complications Associated with Central Venous Catheters among HEN-aligned hospitals than non-HEN-aligned hospitals.

Table 4-9—Difference-in-Differences Impact Analyses For Composite "Any" Adverse Event Outcomes												
Measure	Unadjusted							Regression-Adjusted Difference-in-Differences	Number of Observations	Power to Detect a 5 Percent Effect[a]	Power to Detect a 10 Percent Effect[a]	Power to Detect a 20 Percent Effect[a]
	Non-HEN-Aligned Hospitals			HEN-Aligned Hospitals			Difference-in-Differences					
	Pre 2009–2011	Post 2012–2013	Difference	Pre 2009–2011	Post 2012–2013	Difference						
Any general adverse event (falls or pressure ulcers)	5.44	4.63	-0.80 (0.34)*	5.89	5.36	-0.53 (0.17)**	0.27 (0.38)	**0.04 (0.43)**	125,004	9.3	24.7	72.2
Any post procedural adverse event[e]	5.76	4.59	-1.17 (0.50)*	6.79	6.35	-0.44 (0.25)	0.73 (0.56)	**0.99 (0.56)**	58,663	7.5	17.7	54.3
Any hospital acquired infection[f]	2.39	2.08	-0.30 (0.22)	2.31	2.09	-0.22 (0.10)*	0.08 (0.24)	**0.02 (0.25)**	124,933	7.0	15.9	48.5

Source: The Evaluation Contractor's analysis of MPSMS 2009-2013 data.

Note: Difference-in-differences analyses were regression-adjusted. Robust standard errors, clustered at the hospital level, are in parenthesis. Asterisks indicate statistically significant differences or difference-in-differences at the $p < 0.05$ (*) or $p < 0.01$ (**) levels. These analyses included all at-risk patients in MPSMS sample.

[a]"Power to detect a difference," or the statistical power, is the probability of concluding that the program had a statistically significant effect when the true effect was of the specified size. The power calculation is based on actual standard errors from analysis. For example, in the first row, a 5 percent effect on the "any adverse event" rate would be a change of 0.7 percentage points. Given the standard error of 0.53 from the regression model, the Evaluation Contractor would only be able to detect a statistically significant result 25.1 percent of the time if the impact was truly 0.7 percentage points, assuming a two-sided statistical test at the $p < 0.05$ significance level.

[b]Binary composite measure for having one or more of the 19 adverse events listed in Appendix E. The denominators for composite measures include any cases at risk for one or more of the contributing measures.

[c]Number or count of adverse events per 1,000 patients is a composite continuous measure.

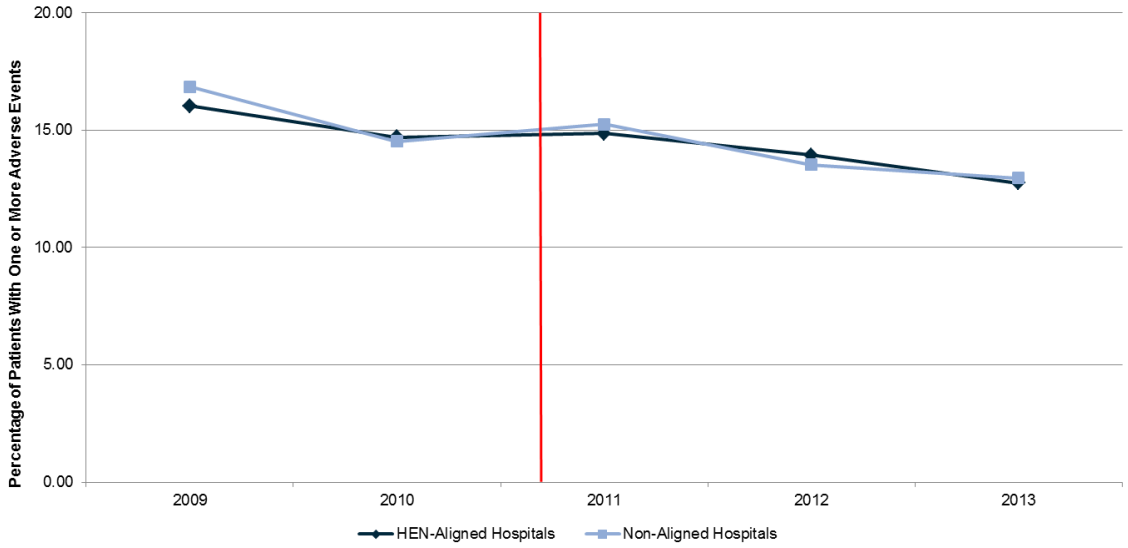
[d]One or more adverse events associated with the following drugs: digoxin, hypoglycemic agents, IV heparin, low molecular weight heparin and Factor Xa inhibitor, or warfarin; or development of antibiotic-associated *Clostridium difficile*.

[e]One or more adverse events associated with the following procedures: femoral artery puncture for catheter angiography, hip joint replacements, knee joint replacements; or one or more of the following adverse events: contrast nephropathy associated with catheter angiography, mechanical complications associated with central venous catheters, postoperative cardiac/non-cardiac arrest events, and postoperative venous thromboembolic events.

[f]One or more of the following HAIs: blood stream infection associated with central venous catheter, catheter-associated urinary tract infections, hospital-acquired vancomycin-resistant enterococcus, postoperative pneumonia, and ventilator associated pneumonia.

Figure 4-39 below provides graphical support for the finding of no impacts associated with HEN alignment. The figure shows regression-adjusted trends for HEN-aligned hospitals and non-HEN-aligned hospitals in the percentage of discharges with any of the 19 adverse events at all. It can be seen that the rates of occurrence of any adverse events were trending downward essentially equally for both non-HEN-aligned and HEN-aligned hospitals.

Figure 4-39—Trends In the Regression-Adjusted Overall Adverse Event Rates, By HEN Alignment



Source: The Evaluation Contractor's analysis of MPSMS 2009–2013 data

A number of robustness checks were conducted on the difference-in-differences estimated impacts of HEN alignment on the composite count of adverse events and the composite "any" adverse event variables, with results that were consistent with the main results, namely that there were no identifiable impacts of HEN alignment. The checks included calculating unadjusted difference-in-differences estimates, and regression models that controlled for hospital fixed, and then random effects. To address the concern that the HENs' work had barely started in 2012, and that including 2012 in the post-period would dilute impacts, the Evaluation Contractor performed an additional check in which hospital discharges occurring in 2012 were dropped from the sample. The main model was then re-run with discharges in 2009 through 2011 (the pre-PfP period) and 2013 alone (the post-PfP period). When the 2012 discharges are removed from the sample, statistical power to detect small effects is poorer than in Table 4-9. For example, the analysis for the main composite measure had just 39.5 percent power to identify a 10 percent effect and 92.3 percent power to identify a 20 percent effect (compared to 73.2 and 99.9 percent, respectively, in the first row of Table 4-9).Results for all of these robustness checks turned out to be roughly similar to the main findings presented in Table 4-9; no results were statistically significant. Full results of these robustness checks are in Appendix E. Adding 2014 data to this analysis may help assess if the difference-in-differences point estimates represented random variability or true effects of the campaign, given the additional statistical power that could be reasonably expected with additional data.

Subgroup Analyses

The Evaluation Contractor considered the possibility that although the main MPSMS impact analyses failed to find evidence of HEN alignment impacts overall, HEN alignment may have had effects for specific subgroups of patients that did not show up in the larger analysis. Several subgroups were investigated at the level of patients (e.g., whether patients were covered by Medicare or not), hospitals (e.g., larger versus smaller or teaching versus non-teaching), and HENs (e.g., those HENs are hospital associations versus those that are system HENs).

The overall findings are that, among the various subgroups analyzed, there was no evidence that HEN alignment had focused impacts among any particular subgroups, or that there were any effects of HENs. The full results of all subgroup analyses are in Appendix E, but among the 155 different comparisons of impacts *within* subgroups (that is, HEN impacts within Medicare-covered patients, or within, hospitals with 200–399 beds, for example), only five estimates (4 percent) were statistically significant at the five percent level; a result that could not be distinguished from random fluctuation. Four of these difference-in-differences estimates were positive, suggesting that the declines in harms in non-HEN-aligned hospitals were actually larger than those in HEN-aligned hospitals. These estimates were thus unfavorable to PfP. One estimate was negative, or favorable to PfP. The four unfavorable and one favorable estimates lacked any clear pattern or possible association with the known work activities of the HENs. One of the unfavorable estimates was for any general adverse events among small HENs (<50 hospitals); two were for any post-procedural events in the Northeast and Southern regions of the U.S.; and one was for any post-procedural events among non-teaching hospitals. Similarly, the one favorable estimate (any ADE among non-Medicare patients) also lacked any clear correlation with HEN activities. This lack of plausible explanations coupled with the fact that one would expect up to 8 statistical tests to be falsely significant among 155 tests due to chance alone, leads the evaluation team to conclude that these estimates do not reflect any true pattern of HEN effects.

Impact Estimates on OB-EEDs and Other Birth Outcomes

The Evaluation Contractor examined available vital statistics data to search for evidence of an impact by one element of the campaign—Strong Start for Mothers and Newborns (Strong Start), which was aimed at reducing OB-EEDs.[4-41] The Strong Start initiative used HENs to encourage their aligned hospitals to adopt the "hard stop" approach, a hospital-level policy prohibiting elective inductions as well as elective primary and repeat Cesarean section (C-section) deliveries before 39 weeks of gestation. This hard stop approach has been shown to be more effective than other approaches at reducing OB-EEDs.[4-42] This analysis examines whether either approach had a measurable impact. In addition to focus on OB-EEDs, HENs worked with their aligned hospitals to reduce other adverse birth outcomes, such as obstetric trauma and injuries to neonates.

[4-41] Strong Start is a broader HHS initiative that includes this PfP component. For more information on the broader initiative, please see http://innovation.cms.gov/initiatives/strong-start/.

[4-42] Clark SL, Frye DR, Meyers JA, et al. Reduction in elective delivery at < 39 weeks of gestation: comparative effectiveness of 3 approaches to change and the impact of neonatal intensive care admission and stillbirth. American Journal of Obstetrics & Gynecology. 2010.

Data

The primary data used for the analyses are the CDC's National Vital Statistics System (NVSS) birth certificate files for calendar years 2009 through 2013 for all 50 states and the District of Columbia (D.C.). The NVSS data contain virtually all births in the U.S. and include indicators for gestational age, method of delivery, whether labor was induced, characteristics of the mother/birth, various perinatal outcomes, and a county identification code signaling the county where a birth occurred. Although the data are collected at the level of the individual birth, information on the hospital where the birth occurred is not available. As a result of this limitation, the Evaluation Contractor compared births in counties with high HEN penetration rates to counties with low HEN penetration rates. That is, the Evaluation Contractor compared births in counties where 90 percent or more of births occurred in HEN-aligned hospitals (the "treatment" group) to births in counties where 50 percent or less of births occurred in HEN-aligned hospitals (the "comparison" group) (see Appendix D). The data set (2009–2013) included a total of 18,265,786 births in 2,133 counties, with 65 percent of births (n = 11,860,740) occurring in treatment counties and 11 percent (n = 2,014,319) occurring in comparison counties (Appendix D).

Methods

Regressions were used to estimate national time trends of six outcome measures that adjusted for patient demographics, risk factors, seasonality, measurement variation, and hospital characteristics (Appendix D contains definitions of the six outcomes studies and a complete list of control variables, definitions, and methods.) The Evaluation Contractor used a difference-in-differences analysis comparing the treatment group to the comparison group in order to estimate the impact of the HEN component of PfP. Details on these methods are included in Appendix D.

Outcome Measures

Six key birth and early delivery-related outcomes were selected as the focus of this analysis. These outcomes were selected on the basis of data availability, and in the case of the two early delivery measures, they reflect the primary focus of the Strong Start Initiative. Four are perinatal health outcomes, including use of assisted ventilation, neonatal intensive care unit (NICU) admission, APGAR score (at 5 minutes), and low-birth weight, and were selected because they have been shown to relate to adverse neonatal outcomes and high economic costs resulting from early (non-medically indicated) elective deliveries.[4-43]

[4-43] Clark et al. 2010; Fowler et al. 2014; Radican-Wald et al. 2014.

Two specific OB-EED measures were also selected, but each was modified slightly compared to measures reported elsewhere. These modifications were as follows:

- **The OB-EED rate**. The analysis used a measure developed by the Medicaid Medical Directors Network (MMDN) rather than the more common Joint Commission's Perinatal Care (PC)-01 measure.[4-44] The PC-01 measure excludes births from the denominator with 39 or more weeks of gestation, as well as births with conditions that possibly justify elective delivery prior to 39 weeks of gestation.[4-45] The Evaluation Contractor elected not to exclude from the denominator births with 39 or more weeks of gestation since these births were also at risk for an OB-EED. That is, because babies born at greater than or equal to 39 weeks of gestation (*not* early deliveries) were also susceptible to being born at 37 or 38 weeks of gestation (if the mother or her provider had pursued an OB-EED), they should be kept in the denominator. While the Evaluation Contractor's measure like the PC-01 measure, excluded births with conditions possibly justifying elective delivery prior to 39 weeks of gestation, the list of conditions available in birth certificates was not as complete as the one used for the PC-01 measure. Please refer to Appendix D for more details.
- **The early induction/early C-section rate**. The simple rates for early inductions or early C-sections include both elective and non-elective early induced deliveries, and early C-sections. The Strong Start initiative's intent was to prevent these procedures only to the extent they were elective (i.e., non-medically indicated). Theoretically, then, any effect of Strong Start on early induction or early C-section should be captured by the measure of the OB-EED rates. However, measuring the overall rate of early inductions/early C-sections may help capture changes in healthcare practice over time, especially if Strong Start changes influenced practices.[4-46] Another factor in including this outcome measure is that the early induction/early C-section rates include data from counties in all states; whereas the OB-EED rate is not available for all states (see Appendix D for more details).

4-44 The Evaluation Contractor modified the MMDN measure by limiting the sample to births occurring in hospitals (instead of all births) because PfP is a hospital-based intervention, and by expanding births to include those not covered by Medicaid. See Fowler et al. 2014.

4-45 The definition of the PC-01 measure is available here: https://manual.jointcommission.org/releases/TJC2013A/MIF0166.html. Some hospitals contend that this does not reflect the positive changes accruing when babies are not delivered early, but continue in the denominator through 39 weeks or later.

4-46 One potential physician response to Strong Start-induced pressure to reduce OB-EEDs is to characterize deliveries that previously would have been characterized as "elective" as "medically necessary." Alternatively, increased attention to OB-EEDs may improve the likelihood that a medically necessary delivery is reported as such. Either shift would result in bias in measuring the true change in OB-EEDs.

Results From Time Trend Analysis of Birth Outcomes

National OB-EED and early induction/early C-section rates have both been declining since 2009. From 2009 to 2013, the OB-EED rate declined from 11.45 to 8.31 per 100 births, or an average yearly decline of 7.71 percent. The early induction/early C-section rate fell from 15.77 percent in 2009 to 12.36 percent in 2013, with an average yearly decline of 5.91 percent per 100 births (Table 4-10).

Table 4-10—Annual National Regression-Adjusted Adverse Outcomes per 100 Births, 2009–2013

Outcome	2009	2010	2011	2012	2013	Average Percent Change[a] Per Year	Number of Births (2009–2013)
Early Induction or C-Section	15.77	14.84	13.65	12.64	12.36	-5.91**	14,378,202
OB-EED	11.45	10.52	9.46	8.52	8.31	-7.71**	10,043,840
APGAR Score 0-6	1.16	1.16	1.09	1.10	1.18	0.51	14,341,322
Assisted Ventilation	2.99	3.02	2.43	2.16	2.12	-8.22**	11,562,808
Admission to NICU	3.45	3.71	3.78	3.97	4.12	4.51**	11,562,808
Low Birth Weight	2.61	2.64	2.63	2.60	2.65	0.31	14,422,867

Source: The Evaluation Contractor's analysis of 2009–2013 NVSS natality files.
Note: All reported rates are multiplied by 100. Rates are regression-adjusted (including county fixed effects) using the variables listed inAppendix E as controls. Statistical significance code: $^{**}p < 0.01$.
[a]Average Yearly Change is the average Growth Rate (r) expressed in percentage points per year, defined as $Y_{2013} = Y_{2009}\,(1+r)^4$.

Despite the decline in both measures of early deliveries, only one of the measures focused on indicators of newborn health outcomes showed a significant decrease–the rate of births requiring assisted ventilation. Low APGAR scores (0-6) were stable over the study period, varying between 1.16 to 1.18 per 100 births, with rates declining during the pre-PfP period and then increasing in the post-PfP period. The percentage of newborns on assisted ventilation declined between 2009 and 2013, despite a rise to 3.02 in 2010 followed by a large decline between 2010 and 2012. In contrast, NICU admissions per 100 births increased slightly, but steadily from 3.45 in 2009 to 4.12 in 2013, although this may reflect the increased availability of NICU units nationwide.[4-47] Low birth weight rates were stable, with an average yearly change near zero over this period.

Since the Strong Start initiative focused on births paid for by Medicaid, it is important to understand how results may have been different for this important subgroup, which constituted roughly 40 percent of the analyzed births.[4-48] Information on insurance source is only available for the last 2 years, and shows that adjusted rates for the six key outcomes tend to be worse for Medicaid-covered births than for non-Medicaid-covered births. Both early induction/early C-section rates, and OB-EED rates increased for Medicaid-covered births but declined among non-Medicaid births between 2012 and 2013. The rates for low APGAR scores, NICU admissions, and low birth weight increased among births for which the primary payer was Medicaid and had a general tendency to also increase for other births. The rate for assisted ventilation declined for both Medicaid-covered births and those births for which the primary payer was not Medicaid. However, assisted ventilation rates increased for births for which the primary payer was unknown or missing.

4-47 Freeman S. Capacity and Utilization in Health Care: The Effect of Empty Beds on Neonatal Intensive Care Admission, Social Science Research Network, March 2015. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2456305. Accessed April 10, 2015.
4-48 The Evaluation Contractor says "roughly" because the primary payers of some births are unknown.

Table 4-11—Adjusted-Rates of Selected Birth Outcomes by Primary Payer, 2012–2013						
Outcome	Primary Payer is Medicaid		Primary Payer is Not Medicaid		Primary Payer is Unknown/Missing	
	2012 (n)	2013 (n)	2012 (n)	2013 (n)	2012 (n)	2013 (n)
Early induction/Early C-section	13.14 (1,064,878)	13.20 (1,095,564)	12.61 (1,405,327)	12.37 (1,439,513)	13.17 (386,813)	12.56 (323,679)
OB-EED	9.15 (924,392)	9.16 (945,895)	8.48 (1,207,461)	8.23 (1,232,492)	7.88 (19,971)	7.61 (19,631)
APGAR score 0-6	1.28 (1,062,993)	1.35 (1,093,731)	1.08 (1,401,470)	1.18 (1,435,678)	0.70 (386,559)	0.74 (323,511)
Assisted ventilation	2.28 (1,063,353)	2.21 (1,094,139)	2.13 (1,402,970)	2.11 (1,437,466)	2.05 (23,041)	3.17 (22,781)
Admission to NICU	4.32 (1,063,353)	4.46 (1,094,139)	3.60 (1,402,970)	3.70 (1,437,466)	4.60 (23,041)	5.65 (22,781)
Low birth weight	3.34 (1,066,120)	3.41 (1,096,545)	2.05 (1,407,173)	2.09 (1,440,665)	2.84 (391,865)	2.77 (326,009)

Source: The CDC NVSS natality data files for all counties nationwide, calendar years 2012–2013.

Impacts of the PfP Strong Start HEN Activities

The analytic results indicate that overall improvement in the six OB-EED measures as a whole was the same, or slightly worse, in the counties with a high proportion of births occurring in HEN-aligned hospitals. OB-EED measure rates were significantly higher for non-HEN-aligned counties than for HEN-aligned counties during the pre-PfP period, but were lower during the post-PfP period. The difference-in-differences estimate for the OB-EED measures was 0.84 percentage points and was statistically significant ($p < 0.01$). None of the difference-in-differences estimates for the other five measures were statistically significant.

Since there was no discernable impact on measures of OB-EEDs, the Evaluation Contractor did not expect that differences in health outcomes would be found. Low APGAR scores were stable over the period, while assisted ventilation trends decreased significantly over the period for both HEN-aligned and non-HEN-aligned counties.

However, NICU admissions during the study period had counter-intuitive results. Most of the observed increase in national NICU admissions during the post-PfP period originated from HEN-aligned counties, with a rate increase from 3.92 in 2012 to 4.13 in 2013, whereas the rate in non-HEN-aligned counties increased only slightly from 3.74 to 3.76 during the same period; however, these changes were not statistically significant. Trends for the percentage of newborns with low birth weight were stable, with no statistically significant change in trends or differences between the groups.

Table 4-12—Regression-Adjusted Rates and Difference-In-Differences (Pre- vs. Post-PfP) Estimates for Six Key Outcomes and HEN-Aligned and Non-HEN-Aligned Counties, 2009–2013

Outcome	All HEN-Aligned Counties					All Non-HEN-Aligned Counties					Difference-in-Differences	Number of Births
	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013		
Early Induction/Early C-Section	16.24 (0.12)	15.22 (0.08)	14.18 (0.05)	13.19 (0.07)	12.81 (0.09)	16.19 (0.31)	15.78 (0.25)	14.18 (0.05)	12.73 (0.21)	12.65 (0.28)	0.41 (0.33)	10,964,575
OB-EED	11.74 (0.13)	10.76 (0.09)	9.95 (0.05)	9.13 (0.07)	8.79 (0.09)	11.96 (0.41)	11.52 (0.23)	9.95 (0.05)	8.39 (0.21)	8.29 (0.22)	0.84** (0.33)	7,413,349
APGAR 0-6	1.29 (0.04)	1.25 (0.02)	1.15 (0.02)	1.13 (0.02)	1.19 (0.03)	1.28 (0.06)	1.26 (0.06)	1.15 (0.02)	1.17 (0.05)	1.29 (0.09)	-0.07 (0.08)	10,933,136
Assisted Ventilation	3.15 (0.24)	3.06 (0.13)	2.65 (0.10)	2.3 (0.11)	2.27 (0.12)	3.7 (0.87)	4.18 (1.27)	2.65 (0.10)	2.84 (0.19)	2.65 (0.20)	0.00 (0.48)	8,553,285
NICU Admissions	3.41 (0.06)	3.71 (0.07)	3.7 (0.04)	3.92 (0.05)	4.13 (0.06)	3.3 (0.21)	3.32 (0.23)	3.7 (0.04)	3.74 (0.18)	3.76 (0.23)	0.15 (0.19)	8,553,285
Low Birth Weight	2.73 (0.02)	2.76 (0.02)	2.73 (0.01)	2.7 (0.01)	2.73 (0.02)	2.78 (0.05)	2.7 (0.06)	2.73 (0.01)	2.61 (0.05)	2.73 (0.06)	0.02 (0.05)	11,000,378

Source: The CDC NVSS natality data files and linked birth/infant death files for all counties nationwide for calendar years 2009–2013.
Statistical significance code: $^{*}p < 0.05$, $^{**}p < 0.01$. Clustered standard errors are shown in parentheses.

Finally, a subgroup analysis examined whether there were differences between the estimated impact of HEN alignment for rural and urban counties (Appendix E). There were no consistent relevant patterns—OB-EEDs were 1.12 percentage points higher for HEN-aligned counties where all births occurred in hospitals located in urban areas ($p < 0.01$) and APGAR scores were 0.36 percentage points lower for HEN-aligned counties where some births were in hospitals located in urban areas ($p < 0.01$).

5. Harm Among Different Subgroups of Hospital Engagement Networks (HENs) and Hospitals

In this chapter, the Evaluation Contractor presents a variety of analyses of data at the HEN-level or HEN-aligned hospital-level. HEN-submitted measure rates and data collected by means of surveys to the HENs and HEN-aligned hospitals are examined by means of statistical process control (SPC) charts, dose-response analyses, cluster analyses, interrupted time series (ITS) analyses, and difference-in-differences analyses. Finally, several examples of notable reductions in patient harms made by different HENs illustrate the broad range of successes achieved by the participants in PfP.

HEN-Level SPC Summary

The Evaluation Contractor used SPC charts to attempt to identify changes in measure results during the Partnership for Patients (PfP) campaign. Briefly, this methodology permits an assessment of whether processes generating changes in Hospital Engagement Network (HEN) data trends show evidence of patient harms being changed in a non-random manner. Two elements of change are considered—whether there is a change in the level of harm, or the specific measure rate for any given time period, and whether there is a change in the trend line, a "bending" of the line toward more (or less) improvement. Results were interpreted as showing evidence for improvement or worsening if the SPC chart showed a shift in the center line, or a change in trend identifiable in six or more consecutive data points, depending on the direction.

The data used in these analyses cover a variety of periods depending on the measure. In many cases, the baseline data were provided for the period before the HEN contracts on the PfP campaign were initiated (e.g., 2010 or 2011). In other cases, the data begin well into the campaign as HENs recruited additional hospitals to their networks and reported data for new cohorts. Data were provided for most measures through Q2 or Q3 2014, but some measures were reported for shorter time frames. The data provided by the HENs were generally all-payer data rather than Medicare or Medicaid patients only. Additionally, the data were typically submitted for only a subset of hospitals working with the HENs. Therefore, some measures are more representative of the HEN performance than others.

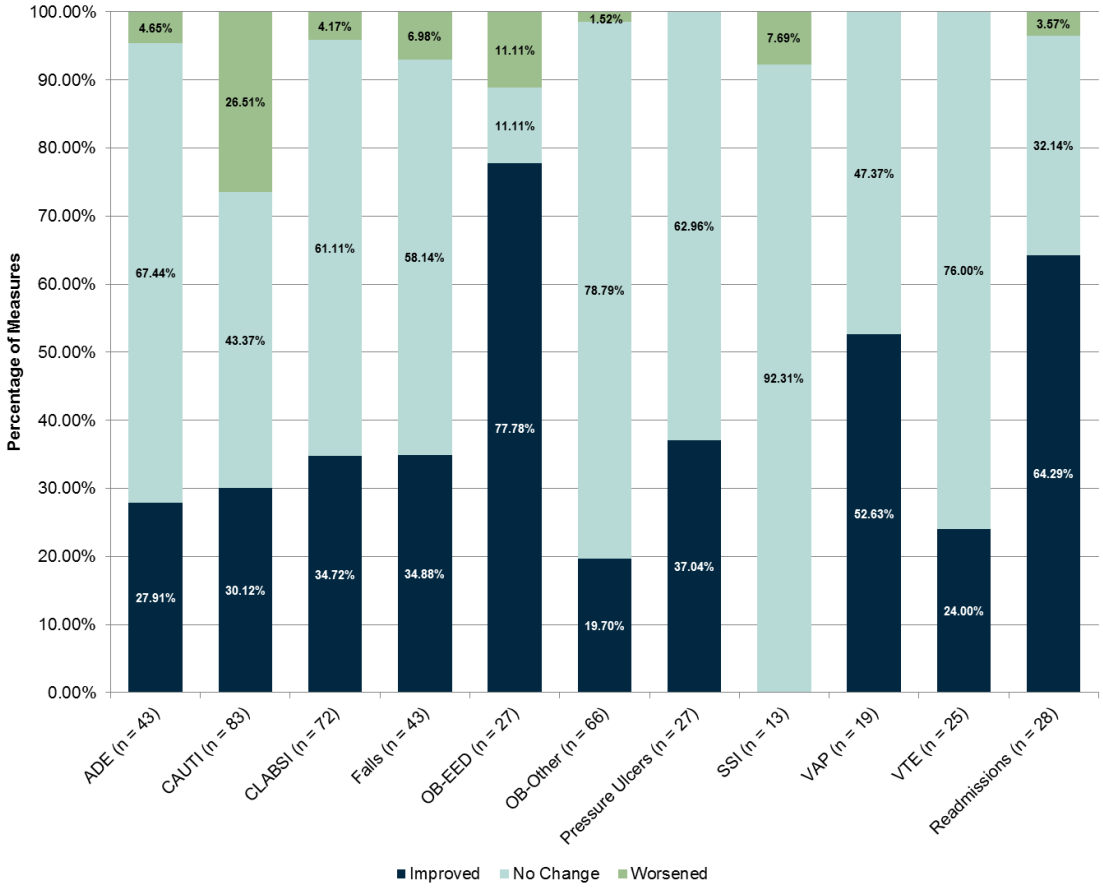
The following figures summarize the evidence for changes in harms based on the SPC charts developed using HEN-submitted data on common measures, which are similar to the national leading indicator measures and have been reported by multiple HENs with comparable specifications. The Evaluation Contractor submitted the HEN-level SPC charts to the Centers for Medicare & Medicaid Services (CMS) via PowerPoint presentations beginning in May 2015.[5-1,5-2,5-3,5-4,5-5,5-6,5-7,5-8,5-9,5-10,5-11,5-12,5-13,5-14,5-15] The full methodology for the development of the SPC charts can be found in Appendix D. A full list of the common measures can be found in Appendix E, identifying the HENs that reported on each measure and their associated finding of improvement, worsening, or no change.

When looking at the HEN-measure combinations, 446 of the outcome measures submitted by the HENs were classified into 37 common measure categories. Of these 446 measures, a majority showed evidence for no change (57.17 percent), followed by measures showing evidence for improvement (34.75 percent). Those measures with evidence for worsening made up the smallest portion of the total measures (8.07 percent).

Figure 5-1 provides a summary of evidence for changes in harms, by adverse event area (AEA). Obstetrical early elective delivery (OB-EED) and readmissions showed the greatest percentage of measures with evidence for improvement with 77.78 percent and 64.29 percent of measures, respectively. Conversely, the AEA with the greatest percentage of measures with evidence for worsening is catheter-associated urinary tract infections (CAUTI) (26.51 percent).

5-1 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Adverse Drug Events. Submitted June 19, 2015.
5-2 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Catheter-Associated Urinary Tract Infection. Submitted July 21, 2015.
5-3 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Catheter-Associated Urinary Tract Infection Standardized Infection Ratio. Submitted May 21, 2015.
5-4 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Central Line Associated Bloodstream Infection. Submitted July 21, 2015.
5-5 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Central Line Associated Bloodstream Infection Standardized Infection Ratio. Submitted May 21, 2015.
5-6 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Falls. Submitted June 12, 2015.
5-7 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Obstetrical Early Elective Delivery. Submitted May 15, 2015.
5-8 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Other Obstetrical Adverse Events Patient Safety Indicator-17. Submitted June 5, 2015.
5-9 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Other Obstetrical Adverse Events Patient Safety Indicator-18. Submitted June 5, 2015.
5-10 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Other Obstetrical Adverse Events Patient Safety Indicator-19. Submitted June 5, 2015.
5-11 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Pressure Ulcers. Submitted May 29, 2015.
5-12 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Readmissions. Submitted May 22, 2015.
5-13 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Surgical Site Infections. Submitted May 29, 2015.
5-14 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Ventilator-Associated Pneumonia. Submitted June 12, 2015.
5-15 Health Services Advisory Group, Inc. Hospital Engagement Network Statistical Process Control Charts: Venous Thromboembolism. Submitted June 12, 2015.

Figure 5-1—Percentage of Common Measures Improving, Worsening, or Not Changing, by AEA



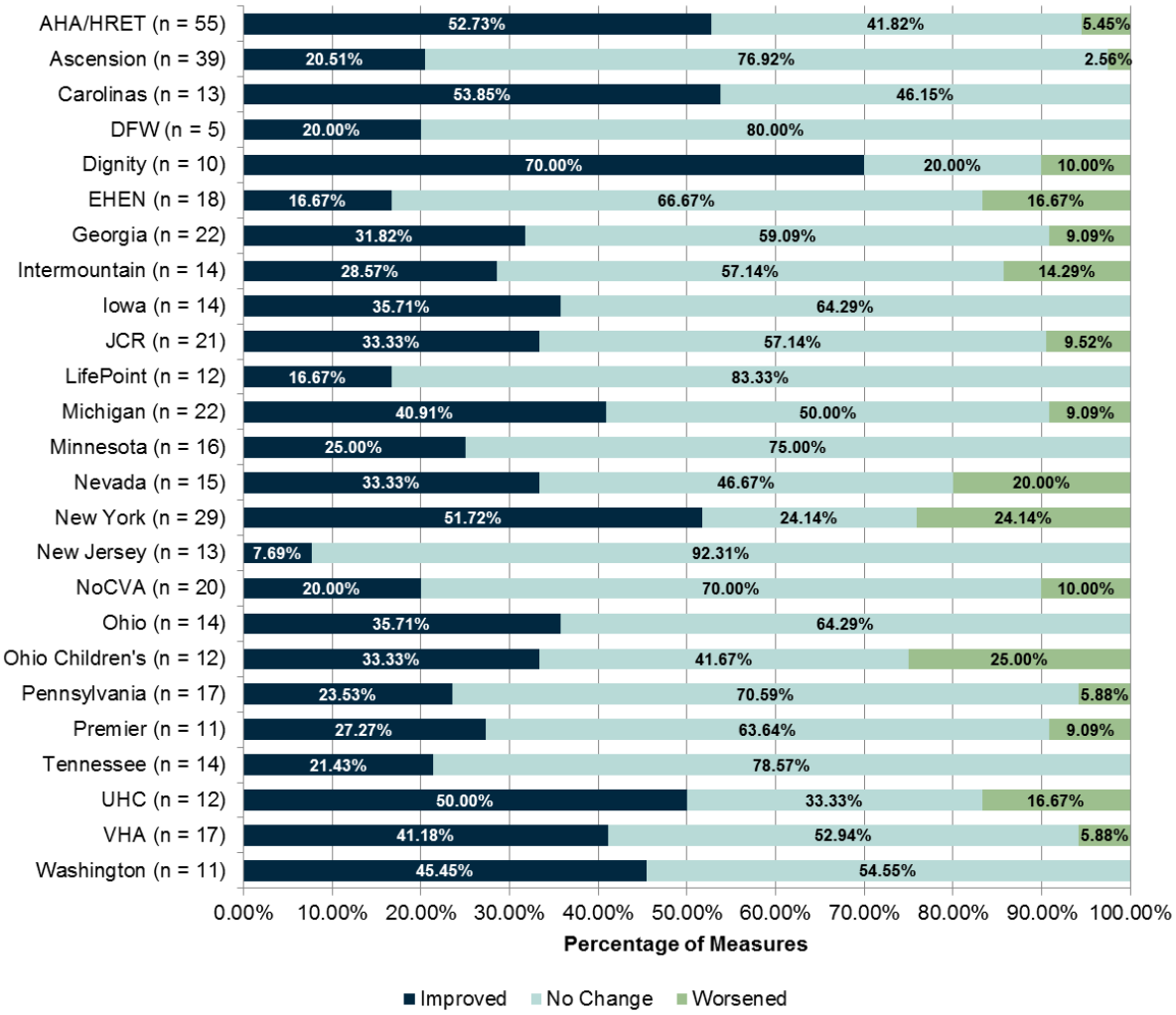
Source: The Evaluation Contractor's analysis of HEN-submitted monthly reports, November 2014.

At the HEN level, the number of HEN-submitted common measures for which an SPC chart could be developed ranged from a low of zero measures (TCQPS) to a high of 55 measures (AHA/HRET).[5-16] Five HENs had at least 50.00 percent of common measures showing evidence for improvement: Dignity (70.00 percent), Carolinas (53.85 percent), AHA/HRET (52.73 percent), New York (51.72 percent), and UHC (50.00 percent). Conversely, three HENs had at least 20 percent of common measures showing evidence for worsening: Nevada (20.00 percent), New York (24.14 percent), and Ohio Children's (25.00 percent). Nine HENs (Carolinas, DFW, Iowa, LifePoint, Minnesota, New Jersey, Ohio, Tennessee, and Washington) did not have any worsening measures.

5-16 TCQPS generally reported only a single pre and post value for each common measure. Given this, there were not enough data points (less than eight data points in the series) available to develop an SPC chart for the HEN's common measures.

Figure 5-2 provides the summary of evidence for changes in harms, by HEN.

Figure 5-2—Percentage of Common Measures Improving, Worsening, or Not Changing, by HEN



Source: The Evaluation Contractor's analysis of HEN-submitted monthly reports, November 2014.
Note: TCQPS generally reported a single pre and post value for each measure. Given this, there were not enough data points (fewer than eight data points in the series) available to develop an SPC chart for the HEN's measures.

Relationship Between Level of Hospital Participation in HEN Activities and Outcome Trends

In their role as one of the three "engines" of the PfP campaign, the HENs developed partnerships, initiatives, and resources to help aligned hospitals reduce adverse events in each target area. Hospitals were free to choose how much to participate in HEN partnerships and initiatives and whether to adopt, adapt, or ignore HEN-provided resources; they could tailor their participation to suit their perceived local needs and circumstances. As a result, there was considerable variation in hospital participation overall and at the level of the HEN and the AEA. This variation makes it possible to ask the research question addressed by this report section: "Were higher levels of hospital participation associated with larger or faster reductions in adverse events?"

Reductions in adverse events were assessed using three metrics:

- The percentage of PfP outcome measures that met campaign harm reduction goals
- The average percentage improvement in PfP outcome measures
- The time required to meet campaign goals

With regard to campaign goals, two sets were examined: the 17.6 percent/10 percent goals of a 17.6 percent reduction in inpatient adverse events and a 10 percent reduction in readmissions, and the 40 percent/20 percent goals of a 40 percent reduction in inpatient adverse events and a 20 percent reduction in readmissions. Goals and improvement for PfP outcome measures were calculated relative to the rates observed during an initial baseline period of at least 3 months. Whether measures met goal and how much they improved was determined based on the rates for the final 3 months in which the measures were observed. Once it was established whether each individual measure met goal, the percentage of PfP outcome measures that met campaign goals was calculated for each AEA within each HEN. Likewise, the percentage change in each PfP outcome measure was determined, and these percentages were averaged for each AEA within each HEN. Time to goal was measured in months, beginning with the first monthly or quarterly observation after the baseline period. In contrast to the other improvement metrics, the unit of analysis for time to goal was the individual measure. All three improvement metrics were calculated from data reported by the HENs to CMS in November 2014.[5-17]

Hospital participation was measured using the Survey of Hospital Participation in Patient Safety Activities. Responses were obtained with HEN assistance in early 2015 from PfP-aligned hospitals.[5-18] For each of the 11 AEAs, hospital respondents were asked whether they received each of the following from their HENs:

- Skills training
- Value-added networking with other hospitals
- Virtual consultation or coaching
- On-site visits
- Feedback on patient safety performance data
- Other education and resources

[5-17] Measures from non-HEN sources were included in the analyses presented in this section; these measures included: 1 CMS readmissions measure; 2 Medicare-source measures regarding pressure ulcers and venous thromboembolisms; 8 National Database of Nursing Quality Indicators® (NDNQI®) measures regarding CAUTI, central line-associated blood stream infection (CLABSI), falls, pressure ulcers, and ventilator-associated pneumonia (VAP); and 6 National Healthcare Safety Network (NHSN) measures regarding CAUTI and CLABSI standardized infection ratios (SIRs) and device utilization ratios, as well as surgical site infection (SSI)-Colon and SSI-Abdominal Hysterectomy SIRs. NDNQI® is a registered trademark of the American Nurses Association (ANA). NDNQI® data were supplied by ANA. The ANA disclaims responsibility for any analyses, interpretations, or conclusions.

[5-18] See Appendix C for additional details on the survey instrument and survey administration.

The percentage of eligible hospitals participating in each patient safety activity was calculated for each HEN and AEA.[5-19] These percentages were used in all analyses as metrics of hospital participation.

The relationships between the three improvement metrics and the metrics of hospital participation were examined using regression methods. Each improvement metric was regressed on hospital participation in each of the six types of activities. HEN and AEA indicators were included in regression models to control for unmeasured factors related to the outcomes. For the percentage of measures that met campaign harm reduction goals and average percentage improvement in measures, observations were weighted by the total number of measures in the HEN and AEA to prevent averages based on very few measures from having an undue influence on results. Clustered standard errors were computed to account for the correlation among measures reported by the same HENs.

To assess whether higher hospital participation in patient safety activities was associated with reductions in the time required to reach campaign goals, Cox proportional hazard regression models were used. For each measure that met goal, time to goal was estimated using a regression method described in Appendix D. The same predictors and controls were used in these models as in the models for the improvement metrics described above. Only measures with at least 8 post-baseline observations were included in the time to goal analyses.

Due to differences in data quality, not all HEN-reported measures were analyzed for this report section. Measures for which at least 60 percent of HEN hospitals reported data in at least 50 percent of the observed periods were considered widely reported. Measures where the hospital count varied by 15 percent or less from the maximum count were considered consistently reported measures. Widely and consistently reported measures are expected to more reliably describe actual trends within HENs compared to measures reported by just a few or a highly variable number of hospitals. Therefore, this section will focus on results for widely reported and consistent measures. Supplemental tables for the remaining measures are presented in Appendix E.

Results

The following analyses depended on the availability of reliable data on hospital participation. Reliable data from the Survey of Hospital Participation in Patient Safety Activities, as determined by survey response rates of 65 percent or higher, were available for 21 out of 30 HENs and cohorts.[5-20] Reliable data were not available for 9 of 30 HENs and cohorts; therefore, the results presented here may not be fully representative of the PfP campaign.[5-21]

5-19 PfP eligible hospitals consist of short stay acute care hospitals, critical access hospitals (CAHs), and children's hospitals. Among PfP-eligible hospitals, AEA-level eligibility was based on a hospital's reported capability to provide specific types of services; e.g., a hospital that did not report having an obstetric unit was not eligible to contribute data regarding the OB-EED or other obstetrical adverse events (OB-Other) AEAs. Hospital eligibility information was obtained from the hospital lists regularly supplied to CMS by the HENs. See Appendix C for additional details.

5-20 The Intermountain, Michigan Cohort 2, NoCVA, and Ohio Children's (Cohorts 1 and 2) HENs did not respond to the survey; and the DFW, Pennsylvania, TCQPS, and Tennessee HENs had response rates below the 65 percent threshold. Both sets of HENs were excluded from analyses.

5-21 In addition, due to steps taken by HENs to protect the confidentiality of aligned hospitals, five of the HENs that met or exceeded the minimum required response rate did not contribute measures for analysis in specific AEAs because of uncertainty over their eligibility to do so. The eligibility of hospitals aligned with EHEN, LifePoint, and New York to report measures concerning the CLABSI AEA were unknown; and the eligibility of hospitals aligned with Minnesota and Ohio to report measures concerning the CLABSI, OB-EED, OB-Other, SSI, VAP/VAE AEAs were also unknown. Therefore, measures from these HENs in these AEAs were excluded from the analyses.

As shown in Table 5-1, the average level of hospital participation in patient safety activities ranged from a low of 31.77 percent for on-site visits to a high of 71.87 percent for feedback on patient safety performance data. Despite the wide gap between these two extremes, the average level of hospital participation for the other activity types fell in a narrow range between 45.67 percent (skills training) and 54.39 percent (other education and resources).

Table 5-1—Descriptive Statistics: Hospital Participation in Patient Safety Activities and Widely-Reported PfP Outcome Measures

Outcome/Activity	Average[a]	Standard Deviation	Minimum	Maximum
Percentage of Hospitals Participating in HEN Activities				
Skills Training	45.67	38.69	6.56	92.31
Value-Added Networking with Other Hospitals	55.65	41.66	0.00	100.00
Virtual Consultation or Coaching	50.33	43.57	0.00	95.83
On-Site Visits	31.77	52.93	0.00	95.65
Feedback on Patient Safety Performance Data	71.87	37.41	11.11	100.00
Other Education and Resources	54.39	34.68	16.67	95.83
Improvement Metrics				
Percent Meeting 17.6 Percent/10 Percent Goals	38.83	59.51	0.00	100.00
Percent Meeting 40 Percent/20 Percent Goals	15.25	43.84	0.00	100.00
Average Percent Change in Measures	-9.39	55.80	-98.47	383.26
Average Months to 17.6 Percent/10 Percent Goal (n = 180)[b]	12.04	9.28	1.00	33.00
Average Months to 40 Percent/20 Percent Goal (n = 70)[b]	13.74	10.19	1.00	32.00

Source: Evaluation Contractor's Survey of Hospital Participation in Patient Safety Activities, January-March 2015; HSAG's analysis of HEN monthly reports, November 2014.

[a]There were 175 HEN/AEA combinations representing 564 widely reported measures.

[b]There were 180 measures meeting 17.6 percent/10 percent goals and 70 measures meeting 40 percent/20 percent goals, each with at least 8 post-baseline monthly or quarterly observations.

Regarding the improvement metrics, on average 38.83 percent of PfP outcome measures met the 17.6 percent/10 percent goals. As would be expected, the average percentage of measures meeting the more rigorous 40 percent/20 percent goals was much lower: 15.25 percent. Individual HEN/AEAs results varied considerably, ranging from 0 percent meeting goals to 100 percent meeting goals. Averages also varied considerably by AEA; e.g., for 17.6 percent/10 percent goals, the average proportion of measures within a HEN/AEA meeting the 17.6 percent/10 percent goals was 18.18 percent for readmissions but 86.67 percent for OB-EED. See supplemental tables in Appendix E for more details.

On average, the HEN/AEA-level change in measures between the baseline period and the end of the post-baseline period was -9.39 percentage points, indicating a modest reduction in harms. The standard deviation of this outcome was 55.80; however, more than six times larger than the mean, pointing to a high level of variation across HENs and AEAs. Substantial variation can also be seen in the range of this outcome, from -98.47 to 383.26—that is, from a near-100 percent decline in harms to a near-quadrupling of harms.

As evidenced by the results reported above regarding the average percentage of measures meeting goals, most measures did not meet campaign goals. Of those that did meet goals, the average time elapsed between the first post-baseline measurement and reaching the 17.6 percent/10 percent goals was 12.04 months. Surprisingly, for the relatively small fraction of measures that met the 40 percent/20 percent goals, the average elapsed time was 13.74 months, barely 2 months longer. There was considerable variation in time to goal by AEA, however; e.g., for the 17.6 percent/10 percent goals, the average elapsed time ranged from a mere 1.67 months for OB-EED to a much-longer 17.88 months for readmissions.

Hospital Participation in HEN Activities Compared to Average Percentage of Measures Meeting Goal

If hospital participation in patient safety activities was more effective in helping hospitals to reduce harms, an association between higher levels of participation and higher proportions of measures meeting goal within each HEN/AEA would be expected. Using regression methods to control for other sources of variation in outcomes, there is limited evidence of a relationship between increased levels of participation in on-site visits and the percentage of measures meeting goals (see Table 5-2). On one hand, a 10 percent increase in the proportion of hospitals participating in on-site visits is associated with a 6.7 percentage point decrease ($p < 0.01$) in the average percentage of measures in a HEN/AEA meeting the 17.6 percent/10 percent goals.[5-22] This association may be spurious, however; a hospital may be more likely to request on-site visits when it is having difficulties reducing harms in an AEA, so that the use of on-site visits and the lower proportion of measures meeting goal are both related to difficulties in reducing harms encountered by hospitals.

Table 5-2—Regression Results: Association Between Hospital Participation in Patient Safety Activities and Percent of Measures Meeting 17.6 Percent /10 Percent Goals

	Percent of Measures in HEN-AEA Meeting 17.6 Percent/10 Percent Goals		Percent of Measures in HEN-AEA Meeting 40 Percent/20 Percent Goals	
Percentage of Hospitals Participating in HEN Activities	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
Skills Training	0.15	0.19	0.01	0.10
Value-added Networking with Other Hospitals	0.05	0.17	0.21	0.11
Virtual Consultation or Coaching	-0.10	0.19	0.04	0.13
On-Site Visits	-0.67	0.24*	0.12	0.18
Feedback on Patient Safety Performance Data	0.46	0.23	0.11	0.14
Other Education and Resources	0.41	0.23	0.46	0.13**

Source: Evaluation Contractor's Survey of Hospital Participation in Patient Safety Activities, January-March 2015; HSAG's analysis of HEN monthly reports, November 2014.

Notes: Each row represents the results of a separate regression analysis with vectors of HEN and AEA variables to control for unmeasured heterogeneity. Analyses include only widely and consistently reported measures. Generalized least squares (GLS) regressions were computed with standard errors adjusted to compensate for clustering of observations within HENs and were weighted by the number of widely and consistently reported PfP outcome measures in each HEN/AEA. Influential observations were deleted based on regression diagnostics; N varies from 165 to 172.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5-22 The 6.7 percent decrease is calculated from the coefficient -0.67, which represents the change in the percentage of measures achieving goal for a 1 percent increase in the percentage of hospitals participating in on-site visits. Thus, -0.67*10 = 6.7 percent.

On the other hand, a 10 percent increase in the proportion of hospitals participating in other education and resources is associated with a 4.6 percentage point increase ($p < 0.01$) in the average percentage of measures meeting 40 percent/20 percent goals.

No other results were statistically significant. This pattern of isolated statistically significant results holds across all analyses presented in this section.

Hospital Participation in HEN Activities Compared to Average Percentage Change in Measures

As noted above, the average percentage change in measures across all HENs and AEAs was slightly less than - 10 percent, representing a modest reduction in harms. It is expected that the following regression analysis will show that increases in hospital participation are associated with decreases in harms.

The regression results in Table 5-3 provide some evidence supporting this expectation; a 10 percent increase in hospital participation in feedback on patient safety performance data is associated with a 3.2 percent average decrease in harms ($p = 0.04$), and a 10 percentage point increase in hospital participation in other education and resources is associated with a 5.0 percentage point average decrease in harms ($p < 0.01$).

Table 5-3—Association Between Hospital Participation in Patient Safety Activities and Average Percentage Improvement in Measure Rates by HEN and AEA

	Average Percentage Improvement in HEN-AEA	
Percent of Hospitals Participating in HEN Activities	**Coefficient**	**Standard Error**
Skills Training	-0.20	0.20
Value-Added Networking with Other Hospitals	-0.21	0.17
Virtual Consultation or Coaching	0.16	0.18
On-Site Visits	0.18	0.12
Feedback on Patient Safety Performance Data	-0.32	0.14*
Other Education and Resources	-0.50	0.15**

Source: Evaluation Contractor's Survey of Hospital Participation in Patient Safety Activities, January-March 2015; HSAG's analysis of HEN monthly reports, November 2014.
Notes: Each row represents the results of a separate regression analysis with vectors of HEN and AEA variables to control for unmeasured heterogeneity. Analyses include only widely and consistently reported measures. GLS regressions were computed with standard errors adjusted to compensate for clustering of observations within HENs and were weighted by the number of widely and consistently reported PfP outcome measures in each HEN/AEA. Influential observations were deleted based on regression diagnostics; N varies from 167 to 173.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

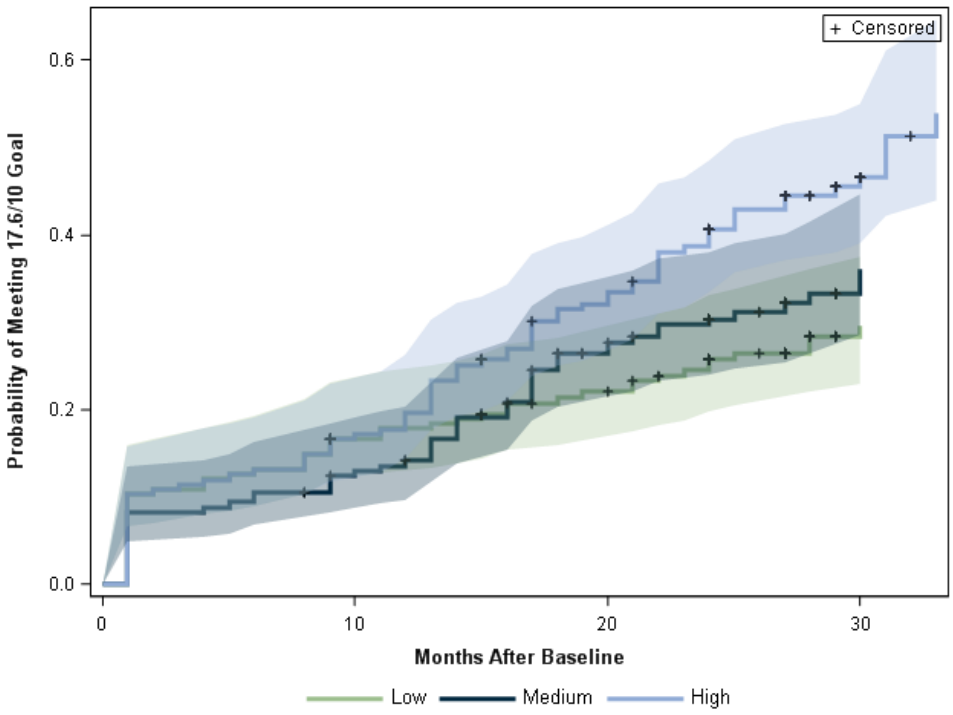
Hospital Participation in HEN Activities Compared to Time to Goal

If higher hospital participation in patient safety activities was more effective in improving performance, then higher levels of participation might be associated with HENs meeting goals more quickly.

Figure 5-3 presents an example of the results for the impact of skills training on the time to reach the 17.6 percent/10 percent goals. It provides suggestive evidence that hospital participation in skills training is associated with shorter times to goal. As shown in Table 5-1, participation in skills training ranged from 6.56 percent to 92.31 percent. This range was divided into three equal-sized portions (tertiles) based on the level of hospital participation. Figure 5-3 clearly shows that the tertile with the highest level of hospital

participation (54 percent or higher) had a higher probability of reaching goal for the entire duration of the PfP campaign. Although overlapping confidence bands preclude drawing firm conclusions about differences between the middle and lower tertiles, by the halfway mark of the campaign the middle does appear to pull away from the lowest tertile. Note that this figure is only descriptive and does not include controls for HEN, AEA, or any other possible confounding variables.

Figure 5-3—Estimated Survival Function, Months to 17.6 Percent/10 Percent Goal, Stratified by Hospital Participation in Skills Training



Source: Evaluation Contractor's Survey of Hospital Participation in Patient Safety Activities, January-March 2015; HSAG's analysis of HEN monthly reports, November 2014.
Notes: The survival functions were estimated using the product-limit method. The shaded areas are Hall-Wellner confidence bands. The distribution by HEN/AEA of hospitals participating in skills training was stratified by tertiles computed after weighting by the number of measures in each HEN/AEA. The "Low" tertile ranged from 0.00 to 36.36 percent (inclusive); the "Medium" tertile ranged from 36.36 to 54.00 percent (inclusive); and the "High" tertile ranged from 54 to 100 percent hospital participation in skills training.

Table 5-4 presents the regression results for each of the types of hospital participation, contrasting time to the two levels of goals.

Table 5-4—Cox Proportional Hazards Regression Analyses: Association Between Hospital Participation in Patient Safety Activities and Number of Months to Reach PfP Goals

	Months to 17.6 Percent/10 Percent Goals		Months to 40 Percent/20 Percent Goals	
HEN Activities	**Hazard Ratio**	**Standard Error**	**Hazard Ratio**	**Standard Error**
Skills Training[a]	1.03	0.01**	0.99	0.02
Skills Training (Over Time)[a]			1.00007	0.00002**
Value-added Networking with Other Hospitals	1.00	0.01	1.00	0.02
Virtual Consultation or Coaching	1.00	0.01	0.99	0.02
On-Site Visits	0.99	0.01	1.00	0.02
Feedback on Patient Safety Performance Data	1.02	0.01	1.01	0.02
Other Education and Resources	1.02	0.01	1.04	0.01**

Source: Evaluation Contractor's Survey of Hospital Participation in Patient Safety Activities, January-March 2015; HSAG's analysis of HEN monthly reports, November 2014.

Notes: Each row except for "Skills Training (Over Time)" represents the results of a separate regression analysis with vectors of HEN and AEA variables to control for unmeasured heterogeneity. Analyses include only widely and consistently reported measures with at least 8 post-baseline monthly or quarterly observations (N = 519).

[a]"Skills Training (Over Time)" is an interaction between the percent of hospitals participating in patient safety activities and the square of the number of post-baseline months to the 40 percent/20 percent goal. It can only be interpreted in conjunction with "Skills training." See text for further elaboration.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For time to 40 percent/20 percent goals, the results are complex. As with the 17.6 percent/10 percent goals, there is evidence of an association between skills training and time to goal. A key difference here is that the relationship between skills training and time to goal changes over the course of the campaign. For example, a 10 percent increase in the proportion of hospitals participating in skills training would be associated with an 8.16 percent lower hazard of meeting goal in the first month of the post-baseline period. However, the association between participation in skills training and time to goal changes over time, so that by the twentieth month of the post-baseline period, the hazard of meeting goal would be 21.78 percent higher.

In addition, an association between hospital participation in other education and resources and time to goal was detected. A 10 percentage point increase in the proportion of hospitals receiving other education and resources was associated with a 4 percent increase in the hazard of meeting the 40 percent/20 percent goals—that is, participation was associated with decreased time to goal.

Hospital Actions and Changes in Policy to Improve Patient Safety

The logic of the PfP campaign was that hospitals would acquire knowledge and tools through participation in HEN activities, they would make changes in practice to implement the best practices they learned about, and this would result in improved patient safety and reductions in harms. This is why, in addition to being asked about participation in patient safety activities, hospital respondents were also asked about whether changes had been made at the hospital level or the unit level as a consequence of participation in PfP patient safety activities and whether such changes had been made for reasons unrelated to PfP participation. The same set of analyses that were conducted for hospital participation in patient safety activities were also conducted for these three measures of hospital changes in practice (see Appendix E for details).

Only two statistically significant results emerged from these analyses: First, there was a relationship between the percentage of hospitals making hospital-wide patient safety improvements and time to meet 40 percent/20 percent goals. This relationship varied with time, so that over most of the PfP campaign a higher percentage of hospitals making changes were associated with longer times to goal. This relationship slowly reversed, so that, by the end of the campaign, a higher percentage of hospitals making changes were associated with shorter times to goal. Second, increases in the percentage of hospitals making changes unrelated to PfP were associated with longer times to meet 40 percent/20 percent goals. Results for non-widely-reported measures were similar.[5-23]

Summary and Discussion

The foregoing set of analyses required the fitting of 30 separate regression models (five improvement metrics individually regressed on each of six distinct types of hospital participation in patient safety activities) and yielded only seven statistically significant results. The most consistent result across the models was an association between the desired outcomes and hospital participation in other resources and education; higher participation was associated with higher percentages of measures meeting 40 percent/20 percent goals, greater average percentage improvements, and shorter times to 40 percent/20 percent goals. Participation in skills training was associated with desired outcomes in two instances; higher participation was associated with shorter times to 17.6 percent/10 percent and 40 percent/20 percent goals. The other two significant results were an association between participation in on-site visits and lower proportions of measures meeting 17.6 percent/10 percent goals on average, and an association between participation in feedback on patient safety performance data and higher average percentage improvements in measures. With one exception, these results only hold for the widely and consistently reported measures; they do not appear in the analyses of the remaining measures (see Appendix E).

The consistency of results for on-site visits and other education and resources goes some distance towards counterbalancing the relative scarcity of statistically significant results. It is not possible to draw firm conclusions about the relationship between hospital participation in patient safety activities and the three improvement metrics considered in this section, but these results are consistent with a weak relationship between at least some of the HEN activities and reductions in patient safety.

The evidence linking the percentage of hospitals making changes to improve patient safety with improved outcomes is even weaker than the evidence linking the percentage of hospitals participating in patient safety

[5-23] Increases in the percentage of hospitals making hospital-wide changes to improve patient safety were associated with longer times to goal. The association between increases in the percentage of hospitals making changes unrelated to participation in PfP activities and times to 40 percent/20 percent goals varied over time; they were initially associated with shorter times to goal, but these quickly turned into longer times to goal.

activities to improved outcomes. This unexpected result raises questions about the reliability of the participation measures with respect to capturing the similarity of activities HENs were providing, the consistency of results exhibited across a diverse set of hospitals and HENs.

While considering this, it is important to take into account the limitations of the data. Survey respondents were not asked about the quality, quantity, or timing of their hospitals' participation in PfP activities. Due to high turnover among hospital management, it is not necessarily the case that reliable answers would have been forthcoming had respondents been asked, because survey respondents new to a hospital would not know about the quality or timing of historical activities. Furthermore, the relationships being assessed in this section are indirect; the data do not permit analysis of the relationship between a hospital's participation and that same hospital's outcomes. The fact that rough measures of hospital participation exhibit minimal evidence of associations with key performance metrics indicates that future quality improvement campaigns should work to develop stronger research designs for identifying relationships.

Relationship Between The Level of Hospital Engagement and Reductions in Adverse Events

In many of the Evaluation Contractor's analyses, engagement is represented by a hospital's alignment with a HEN. However, some non-aligned hospitals worked with a HEN, and not all HEN-aligned hospitals were equally engaged in HEN activities. For this analysis, the Evaluation Contractor considers the effect of a hospital's level of engagement with its HEN as measured by survey data to determine whether there was a "dose-response" relationship, where more engagement in HEN activities is associated with better change in outcomes, as measured by Medicare fee-for-service (FFS) claims data.

Hospitals that self-identified as working on patient safety improvement with a HEN or state hospital association (SHA) subcontracted with the AHA/HRET HEN (SHA) were asked in a survey to identify whether they were fully engaged, moderately engaged, minimally engaged, or not at all engaged in patient safety activities sponsored or led by a HEN for each harm area (2014 Survey on Prevention of Adverse Events and Readmissions, see Appendix C for survey details). All survey respondents that said they had not worked with a HEN or SHA on patient safety improvement during 2012-2013 were considered not at all engaged in HEN activities for each of the relevant adverse event areas.

In this analysis, changes in outcomes among hospitals that self-reported being fully engaged in HEN activities for a particular adverse event area were compared to hospitals that were not engaged or had reported not working with a HEN or SHA ("non-engaged") using difference-in-differences models. Propensity-score reweighting was used to better assure comparability between the fully engaged group and the non-engaged group on factors other than the level of engagement. Similarly, changes in outcomes were compared for hospitals with other levels of engagement (moderate, minimal) versus the non-engaged group to determine if the overall pattern supports a "dose-response" relationship, where more engagement in HEN activities is associated with better change in outcomes.

The outcomes examined in this analysis, as measured in the Medicare claims data, included Medicare FFS 30-day all-cause readmissions and the Agency for Healthcare Research and Quality (AHRQ) patient safety indicators (PSIs) for central venous catheter-related blood stream infections (CRBSI [PSI-07]), pressure ulcers (PSI-03), and venous thromboembolism (VTE [PSI-12]).[5-24]

Across the four areas, there was no evidence that hospitals that were more engaged in HEN activities had better outcomes. The difference-in-differences impact estimates found no statistically significant differences in the change in the readmissions, CRBSI, pressure ulcer, or VTE rates for any level of engagement among HEN-aligned hospitals (Table 5-5).

These results are consistent with those above, which suggest that greater levels of engagement in HEN activities did not result in greater improvements in outcomes. However, it is possible and even likely that some hospitals that were classified as non-engaged actually received benefit through PfP but did not recognize the source of that benefit as PfP, and/or that hospitals reporting other levels of engagement under-reported their true engagement. This could occur due to the robust HEN partnerships with organizations that served broader constituencies. For example, HEN-Quality Improvement Organization (QIO) partnerships were common. When a QIO was executing technical assistance to hospitals that was jointly planned with the HEN and possibly funded under a PfP subcontract, would the participating hospital recognize the QIO as PfP-funded, or just think of them as the QIO? Such a hospital could respond to the survey that they did not work with a HEN or SHA and thus be categorized in the non-engaged group.

Table 5-5—Estimated Impacts of Hospital Engagement in HEN Activities on Medicare Patients' Adverse Event Rates

	Readmissions[a] N= 1,002 Hospitals	CRBSI (PSI-07)[b] N= 930 Hospitals, 4,132,323 Discharges	Pressure Ulcers (PSI-03)[b] N= 1,018 Hospitals, 2,099,631 Discharges	VTE (PSI-12)[b] N= 914 Hospitals, 1,686,262 Discharges
Fully Engaged	-0.91 (0.57)	-0.14 (0.10)	1.13 (0.81)	-0.94 (0.72)
Moderately Engaged	-0.34 (0.68)	0.11 (0.21)	1.69 (0.97)	-1.75 (1.24)
Minimally Engaged	0.31 (0.61)	0.46 (0.52)	0.58 (0.59)	-1.13 (0.90)

Source: Evaluation Contractor's analysis of Medicare inpatient claims data.

Note: The difference-in-differences impact estimates shown in this table represent differences between the changes in outcomes of the HEN-aligned hospitals with full, moderate, or minimal engagement in HEN activities between baseline and 2013 follow-up relative to changes in the non-engaged HEN-aligned and non-HEN-aligned hospitals over the same period. Engagement in HEN activities was reported on the 2014 Survey on Prevention of Adverse Events and Readmissions.

Each row corresponds to a separate difference-in-differences regression using propensity score-based weights. The impacts were estimated by using linear probability regression models. Robust standard errors (SE), clustered by hospital, are reported in parentheses.

*Difference-in-differences treatment-comparison impact estimate significantly different from 0 at the .05 level, two-tailed test.

**Difference-in-differences treatment-comparison impact estimate significantly different from 0 at the .01 level, two-tailed test.

[a]30-day all-cause readmissions per 100 discharges.

[b]Reported as the number of adverse events per 1,000 Medicare beneficiary discharges at risk.

5-24 The definitions for the AHRQ PSI indicators can be found here: http://www.qualityindicators.ahrq.gov/modules/PSI_TechSpec.aspx.

Relationship Between HEN Characteristics and Making Operational Changes Due to Participation in HEN Activities

Using the 2015 Survey of Hospital Participation in Patient Safety Activities, the Evaluation Contractor examined the associations between HEN characteristics (ownership, size, and rural composition) and hospitals' making operational changes made due to participation in HEN activities.

When looking at hospitals' indications of making changes due to participation in HEN activities across three HEN characteristics, only ownership had a significant association with the odds of hospitals making operational changes due to participation in HEN activities ($p < 0.0001$). Hospitals in system-owned HENs were 1.79 times more likely than hospitals in "other" HENs to indicate they made changes due to participation in HEN activities in at least three applicable harm areas. Similarly, hospitals in SHAs were 1.43 times more likely than hospitals in "other" HENs to have made operational changes due to participation in HEN activities in at least three harm areas. "Other" HENs include national membership organizations, region-specific or national collaboratives, and standing organizations functioning as quality improvement resources. Table 5-6 presents the ratios for the odds of making changes due to participation in HEN activities by HEN characteristic.

Table 5-6— Relative Risk Ratios by HEN Subgroup

HEN/SHA Characteristic	Subgroup	Relative Risk Ratio
Ownership	System (vs. Other)	1.79
	SHA (vs. Other)	1.43
Size	<50 Hospitals (vs. 100+ Hospitals)	1.03
	50-99 Hospitals (vs. 100+ Hospitals)	1.03
Rural composition	>30% Rural (vs. 0-30% Rural)	1.05

Source: Survey on Participation in Patient Safety Activities.
Note: N = 2,159 hospitals

Relationship Between HEN Activities and Outcome Trends

ITS Cluster Analysis

The measures used for ITS analysis provide a detailed look into the performance of HEN-level outcome measures. The results of analyses examining the overall change in measures indicate that HEN outcomes showed substantial improvements during the course of the PfP campaign. However, just over 1 in 5 (22.37 percent) of the outcome measures at the HEN level exhibited significant structural breaks toward more improvement, a result indicating that the processes contributing to patient harm rates were changing. The purpose of this analysis is to assess the relationships between HEN activities and the outcomes they intended to reduce.

The typical ITS analysis would use specific information on the timing of implementation of an intervention as a proxy for the timing of expected changes in the outcome. Under circumstances of a single intervention, the Evaluation Contractor can examine the level and trend in the post-intervention trend as compared to the pre-intervention trend for evidence of a program impact. When more than one intervention is implemented, the ITS framework can be used to examine changes in outcome trends associated with the implementation of different program components. However, for multiple intervention analysis, there must be sufficient differences in the timing of component implementations to permit differentiation of the impact of each intervention from the others. In the PfP campaign, although there were large numbers of activities reported by each HEN, many of the HEN activities were implemented simultaneously with other activities, presenting an evaluative challenge when attempting to link any specific activity to a reduction in patient harms.

In data collected in fall 2014, the HENs reported over 5,000 activities aimed at spreading best practices and reducing patient harms among their constituent hospitals (see Appendix B for more details). In any given HEN and AEA, there were multiple activities implemented on an ongoing basis. With outcome measures having as few as 8 observations, and frequently having fewer than 36 observations, the data does not provide enough statistical power to assess the potential impact of each of these interventions at the same time in an ITS model. Additionally, since many activities were implemented at the same time or in close proximity to the timing of other interventions, ITS models that provide evidence of significant structural breaks at specific points in time cannot tie individual activities to the break in the outcome. Given these complications with a traditional ITS approach, and the difficulty of isolating a unique impact of individual activities at the HEN level, the Evaluation Contractor used a more exploratory approach to identify potential associations between the HEN-reported activities and the reductions in patient harms and structural breaks detected in the outcome measures.

Cluster Analysis Methodology

The Evaluation Contractor used a hierarchical cluster analysis to examine the characteristics of HEN-level outcomes, with respect to the nature and timing of activities the HENs were engaged in at the time the data was observed. Cluster analysis techniques were used in an exploratory manner to identify any patterns of exposure to HEN activities experienced by the outcome measures and the outcome performance on two metrics: exhibiting a structural break in the ITS model, and overall improvement. The cluster analysis began by grouping outcome measures according to their similarity in exposures to four different classifications of activities: tool dissemination, education, one-on-one coaching, and leadership transformation. Specifically, for each observation in each outcome, the Evaluation Contractor counted the number of HEN activities reported across four broad classifications of activity. For each observation, the number of activity-months for each type of activity was counted. For example, if an observation for January 2013 was exposed to educational activities, that observation was counted in the total education activity-month count. For data

reported quarterly, the beginning and ending dates of the HEN activities were used to define the number of activity-months to include. The expectation is that greater exposure to activities should be associated with greater reductions in patient harms or greater number of structural breaks identified.

In addition to counting the number of activity-months of exposure to different types of activities for each observation in each outcome, the Evaluation Contractor also calculated the percentage of observations for each outcome that were exposed to a particular type of activity. Thus, if a measure were reported quarterly for 2 years (i.e., 8 quarterly observations) and the measure had been exposed to educational activities in 4 of the 8 quarters, then the percentage of observations exposed to education would be 50 percent.

The hierarchical clustering algorithm used for the analysis was performed once for each of the common measures reported by multiple HENs, when there were at least five measures available for inclusion.[5-25] The methodology began by treating each outcome measure as its own cluster in the analysis and then systematically combined similar clusters one-by-one, minimizing the increase in the variations across clusters. If two measures truly are similar in their exposure characteristics, combining them into a single cluster should not increase the overall variation across all clusters in the exposure characteristics. In the extreme case, two clusters that are identical in their exposure characteristics can be combined with no increase in the variation within the clusters. This process continued until all outcomes observations were combined into a single cluster. At each step in the combinatory algorithm, the percentage of the variation across the outcomes explained by the clusters is calculated. To identify an optimal number of clusters, the Evaluation Contractor used the number of clusters identified that would explain at least 70 percent of the variability across the common measure, using the R-squared statistic.[5-26]

Once the number of clusters was identified, the average percentage improvement across measures and the average number of structural breaks identified by the ITS analysis were calculated, and the results were visually inspected for any patterns indicating potential relationships between HEN activities and the performance of outcome measures. To be clear, hierarchical clustering methods are exploratory in nature. The results should not be interpreted as definitive evidence of a correlation between the HEN-reported activities and outcome measure performance. Rather, the evidence provided by this analysis should be used to focus attention on the broad types of activities that may have been more closely aligned with reductions in patient harms. Evidence of a possible relationship would exist if there were a consistent pattern between the relative magnitudes of the HEN activity exposures across clusters when compared to the relative magnitudes of the outcomes from the ITS and overall change analysis. Further evidence would be needed to identify clear evidence of correlation or causal impact.

[5-25] Clustering algorithms require that there be enough outcomes observed to generate clusters, or groupings, of outcomes with similar characteristics. While there is no commonly agreed upon lower threshold to the number of observations included in the analysis, the Evaluation Contractor did not evaluate any common measures for which there were fewer than five outcome measures to include.

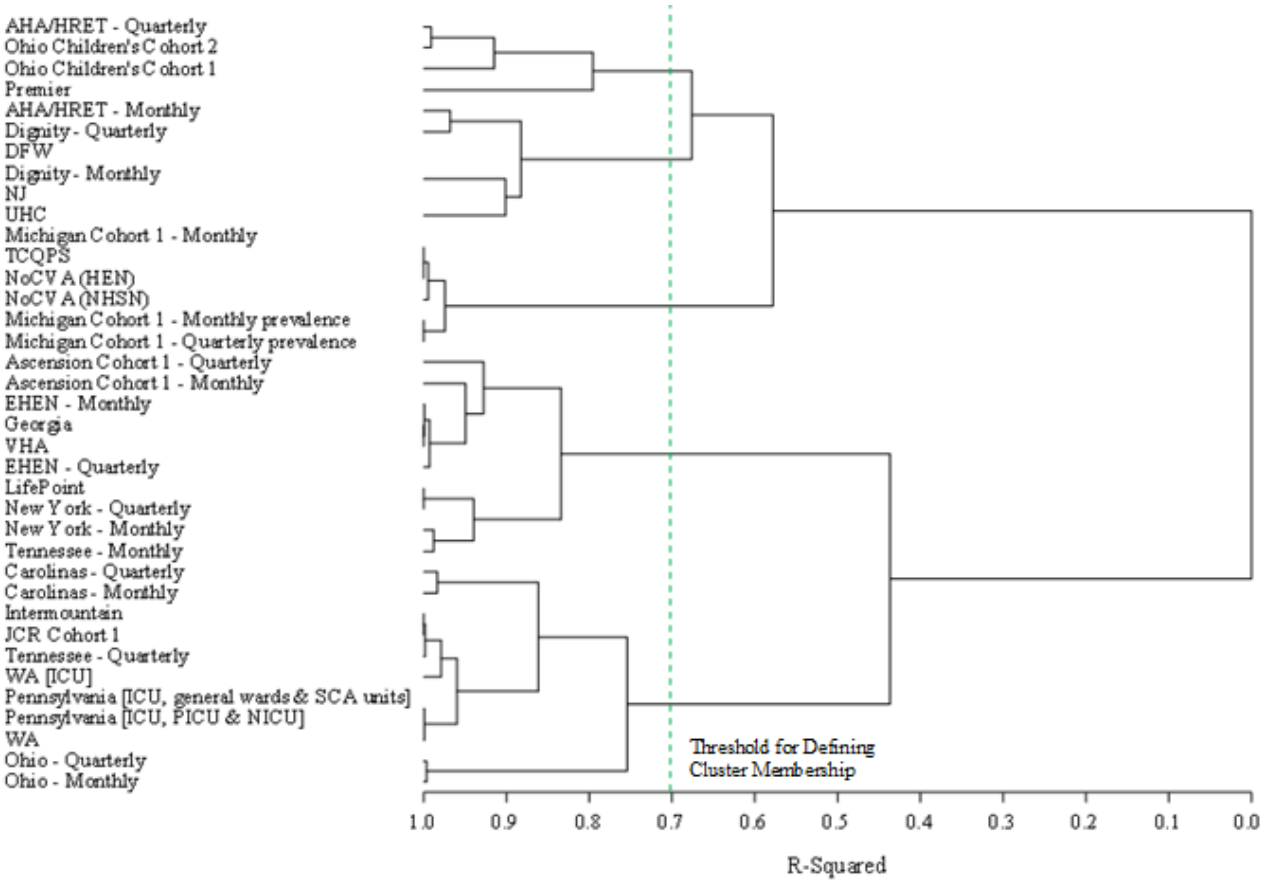
[5-26] Hierarchical clustering methods are exploratory in nature; therefore, the 70 percent threshold does not represent the only threshold that could be used. The 70 percent threshold was chosen to balance the amount of variation in the outcome characteristics with the number of clusters identified. Results could vary if a higher or lower threshold were chosen.

Cluster Analysis Results

The primary results of hierarchical cluster analysis are captured in two pieces of output. First, Figure 5-4 presents a dendogram showing the order in which HEN-level outcomes were combined for one of the common measures examined: the CAUTI device utilization ratio. (A separate dendogram was constructed for each of the common measures.) There were 37 HEN-submitted measures related to CAUTI device utilization ratios.[5-27] The second output is a table presenting the average exposure to different types of HEN activities for each cluster as well as the average number of structural breaks identified in ITS analysis and average reduction in harm.

The results presented in the dendogram indicate that the original 37 CAUTI device utilization measures could be combined into a smaller subset of clusters while simultaneously retaining a relatively high degree of variation captured by cluster membership. At the 70 percent R-squared threshold, five clusters were identified, each representing between 4 and 11 CAUTI device utilization measures. Thus, the cluster memberships alone explain approximately 70 percent of the differences in the HEN activity exposure patterns across all measures, while simultaneously maximizing the similarity of exposure across measures within each cluster.

Figure 5-4—Dendogram of CAUTI Device Utilization Ratio Clustering Algorithm



Source: Evaluation Contractor analysis of HEN Monthly Report Data, November 2014, and HEN timeline data, fall 2014.

5-27 Some HENs appear more than once in the analysis because they reported outcomes for multiple cohorts of hospital participants or because they included CAUTI device utilization measures for multiple units within hospitals (e.g., intensive care unit [ICU], ICU-plus, and hospital-wide).

Having identified the cluster memberships based on HEN activity exposures, Table 5-7 presents the number of HEN outcomes belonging to each cluster, the HEN activity exposure levels for the measures in each cluster, the average number of structural breaks identified in ITS analysis, and the average overall improvement in the outcomes for each cluster. The results show that the cluster with the largest average improvement in CAUTI device utilization was Cluster 3, with a 14.68 percent average improvement in the outcomes. The HENs reporting measures in this cluster were exclusively focused on educational activities, with 22.00 average months of experience, representing 83.07 percent of the observed data points across the series. The results, however, do not point to a direct link between higher exposures to educational activities and improved performance. Cluster 2 and Cluster 4 exhibit comparable exposure to education as observed in Cluster 3, with 21.45 and 22.1 average months, respectively. The percentage of observations exposed to education was only slightly lower than Cluster 3 as well, at 69.29 and 71.74 percent of observed data points, respectively. Yet despite this degree of exposure, the measures in Clusters 2 and Cluster 4 improved by only 9.14 and 7.98 percent, respectively.

The remaining results in Table 5-7 do not indicate any clear patterns of results that suggest a linkage between HEN activities and either the number of structural breaks detected or overall percentage improvement.

Table 5-7—CAUTI Device Utilization Ratio HEN Activity Exposure and Performance Results												
			Average Activity Experience (in Months)				Average Percentage of Data Points with Activity				Outcomes	
Cluster	Number of HENs	Average Number of Hospitals	Tools	Education	Leadership	Coaching	Tools (%)	Education (%)	Leadership (%)	Coaching (%)	Average Number of ITS Breakpoints	Average Percentage Improvement
1	6	74.83	4.33	1.50	1.33	2.67	16.85%	5.56%	15.00%	9.07%	0.00	8.98%
2	11	52.64	15.00	21.45	0.27	19.36	52.20%	69.29%	2.02%	69.22%	0.54	9.14%
3	6	62.67	0.00	22.00	0.00	0.00	0.00%	83.07%	0.00%	0.00%	0.50	14.68%
4	10	64.40	23.40	22.10	17.90	19.70	82.75%	71.74%	55.77%	60.45%	0.50	7.98%
5	4	290	3.25	12.00	19.75	5.25	16.96%	23.92%	30.48%	19.44%	0.50	(8.57%)

Source: Evaluation Contractor analysis of HEN Monthly Report Data, November 2014, and HEN timeline data, fall 2014.

The results presented in Table 5-7 provide one example of the results developed during the hierarchical cluster analysis for a single common outcome measure. The resulting tables for the common measures with at least five outcome observations are included in Appendix E with detailed results. A summary of the results is provided below in Table 5-8. In Table 5-8, the Evaluation Contractor listed results for which the relative levels of HEN activity exposures across clusters are consistent with the relative levels of the outcome performance measures across clusters. For example, if the clusters with the highest or lowest values for the number of tool-months are the same as the clusters with the highest or lowest values of the average number of structural breaks identified by ITS, then "tools" would be listed in the "Exposure Months Related to Structural breaks" column.

Among the common measures, 14 of the 37 did not have at least five outcome measures to be included in the cluster analysis. Of the remaining 23 measures shown in Table 5-8, results consistent with an association between HEN activities and outcome performance were found for only 36 relationships out of 368 (9.78 percent) examined. Thus for the large majority of common measures, there is little evidence linking the exposure of outcome to HEN activities to the final outcome performance.

Table 5-8—Results Summary for Hierarchical Cluster Analysis of Common Measures						
Measure	**Number of Clusters**	**Number of Measures**	**Exposure Months Related to Structural Breaks**	**Exposure Percentage Related to Structural Breaks**	**Exposure Months Related to Overall Improvement**	**Exposure Percentage Related to Overall Improvement**
Adverse Drug Event (ADE)						
Clostridium difficile (*C. difficile*)	3	5	None	None	Tools Leadership Education	Tools Leadership
CAUTI						
CAUTI Device Utilization Ratio	5	37	None	None	None	None
CAUTI per 1,000 Catheter Days (Intensive Care Unit [ICU])	3	7	Tools Education Leadership	Tools Leadership	None	None
CAUTI per 1,000 Catheter Days (ICU-plus)	3	9	Education	None	Leadership	None
Central Line-Associated Bloodstream Infection (CLABSI)						
CLABSI standardized infection ratios (SIR)	3	29	Tools	Tools	Tools	Tools
CLABSI Device Utilization Ratio	4	37	Tools Coaching	Tools Coaching	None	None
CLABSI per 1,000 Central Line Days (ICU)	4	14	None	None	None	None
CLABSI per 1,000 Central Line Days (ICU-plus)	3	9	None	None	None	None
Falls						
Falls per 1,000 Patient Days National Database of Nursing Quality Indicators (NDNQI®)	4	13	None	None	None	None
Falls with Injury per 1,000 Patient Days (NDNQI)	4	14	None	None	None	None

Table 5-8—Results Summary for Hierarchical Cluster Analysis of Common Measures						
Measure	**Number of Clusters**	**Number of Measures**	**Exposure Months Related to Structural Breaks**	**Exposure Percentage Related to Structural Breaks**	**Exposure Months Related to Overall Improvement**	**Exposure Percentage Related to Overall Improvement**
OB-EED						
Perinatal Care (PC)-01 Early Elective Delivery	4	16	None	Education	Leadership	Leadership
Other Obstetrical Adverse Event (OB-Other)						
PSI-17: Injury to Neonate	3	16	None	None	None	None
PSI-18: Obstetric Trauma – Vaginal Delivery with Instrument	4	17	None	None	None	None
PSI-19: Obstetric Trauma – Vaginal Delivery without Instrument	4	19	None	None	None	None
Pressure Ulcers						
PSI-03: Medicare	4	25	None	None	Tools	Tools
PSI-03: All-Payer	4	14	Coaching	None	None	None
Surgical Site Infections (SSI)						
SSI – Abdominal Hysterectomy SIR	5	26	Leadership	Leadership	None	None
SSI – Colon Surgery SIR	6	28	None	None	None	None
Ventilator-Associated Pneumonia (VAP)/Ventilator-Associated Event (VAE)						
VAP per 1,000 Ventilator Days	3	6	None	None	None	None
VTE						
PSI 12: Perioperative PE/DVT (Medicare)	5	25	None	None	None	None
PSI 12: Perioperative PE/DVT (All Payer)	4	16	None	None	None	None

Table 5-8—Results Summary for Hierarchical Cluster Analysis of Common Measures						
Measure	**Number of Clusters**	**Number of Measures**	**Exposure Months Related to Structural Breaks**	**Exposure Percentage Related to Structural Breaks**	**Exposure Months Related to Overall Improvement**	**Exposure Percentage Related to Overall Improvement**
Readmissions						
30-Day All-Cause Medicare Readmissions	4	28	None	None	None	None
30-Day All-Cause All-Payer Readmissions	3	17	Coaching	Education Coaching	Tools Education Leadership	Tools Leadership

Source: Evaluation Contractor analysis of HEN Monthly Report Data, November 2014, and HEN timeline data, fall 2014.

Relationship Between Dosage of HEN Activity and Improvements in Level and/or Trend of Outcomes

The results of ITS analysis indicate that approximately 1 in 5 of the HEN-level outcome measures (22.37 percent) exhibited a structural break toward improvement. Of those that exhibited a break in a direction consistent with a campaign impact, approximately half also exhibited overall improvement (11.78 percent of the 1,265 measures examined with ITS). A natural question to ask is whether or not HEN interventions geared toward reducing patient harms were associated with any observed reductions.

In contrast to the method used in the cluster analysis discussed previously, this analysis presents another exploratory examination of the ITS results, stratifying the model results according to overall levels of HEN activities, regardless of the type of activity. This permits examination of whether the overall level of HEN activity impacted the relative level of improvement in outcomes, and if so, whether the impact was the same over different performance metrics.

Dose-Response Methodology

The analysis proceeded by tabulating the total number of activities reported by each HEN in their HEN activity timelines, for each AEA (see Appendix B and Appendix D for more details). The distributions of activities across HENs in each AEA were examined for evidence of logical cut points for the formation of low, medium, and high activity groupings. Each of the distributions exhibited a substantial degree of positive skew, so that the majority of HEN activity counts were in the low to moderate range, and a few HENs with relatively high levels of reported activities formed a group of high activity HENs. For that reason, HENs with activity counts that fell in the lowest quartile were defined as the low activity group. HENs for which activity counts fell between the 25th percentile and the 50th percentile were classified as the medium activity group. Finally, HENs with activity counts above the 50th percentile were classified as high activity HENs.

Having defined the low, medium, and high activity HENs for each AEA, the Evaluation Contractor examined the ITS results reported by HENs in each activity group. Table 5-9 presents the overall results comparing the three HEN activity groups to different measures of HEN-level outcome.

Results

First, the percentage of HENs' measures that met the 40 percent/20 percent goals, the 17.6 percent/10 percent goals, or that did not improve were compared across levels of HEN activity. One might expect that, as the level of HEN activity increased, so would the percentage of measures meeting goal. However, some of the results were inconsistent with this hypothesis. Although higher levels of HEN activity were associated with higher percentages of measures meeting 40 percent/20 percent goals, this does not appear to hold true for the measures meeting the 17.6 percent/10 percent goals. In that group, improvement levels were similar across all three levels of activity, differing by only 2 percentage points. Also somewhat counterintuitively, the percentage of measures not meeting goal was highest for HENs with medium levels of activity and lowest for HENs with the highest activity levels, with percentages of 57.88 and 52.11 percent, respectively. These results are presented in Table 5-9, along with the other dose stratification results.

Next, the average changes in performance, as measured from baseline to the last three months of each measure, were compared across levels of HEN activity. Among high activity HENs, improving measures changed by a larger percentage (a 40.32 percent reduction) than was seen among HENs with low or medium activity levels (reductions of 34.03 and 34.83 percent respectively). However, at the same time, worsening measures exhibited a larger average increase in rate among high activity HENs (97.10 percent increase) than among low activity HENs (a 62.47 percent average increase in patient harms).

Table 5-9—Stratification of Outcome Metrics by Activity Level and HEN Timeline Data

Result	Level of Activity							
	Low		Medium		High		Total	
	Percent	N	Percent	N	Percent	N	Percent	N
Goal Performance								
Percent of Measures that Met 17.6 Percent/10 Percent Goal	21.98%	153	19.77%	68	21.27%	94	21.26%	315[a]
Percent of Measures that Met 40 Percent/20 Percent Goal	19.68%	137	21.51%	74	25.57%	113	21.86%	324
Percent of Measures that Did Not Meet Goal	54.13%	406	57.88%	202	52.11%	235	54.39%	843
Measures Tested for Overall Change								
Average Change in Rate (Improving)	-34.03%	270	-34.83%	158	-40.32%	187	-36.15%	615
Average Change in Rate (No Sig Change)	-5.78%	334	-6.44%	144	-2.05%	196	-4.84%	674
Average Change in Rate (Worsening)	62.47%	73	84.40%	35	97.10%	52	78.52%	160
Changes in Measure Trends (ITS)								
Percent of Measures with Break Toward Greater Improvement	17.92%	98	26.30%	81	25.43%	104	22.39%	283
Percent of Measures with No Significant Break	67.64%	370	58.44%	180	55.99%	229	61.63%	779
Percent of Measures with Break Toward Less Improvement	14.44%	79	15.26%	47	18.58%	76	15.98%	202

Table 5-9—Stratification of Outcome Metrics by Activity Level and HEN Timeline Data								
Result	Level of Activity							
	Low		Medium		High		Total	
	Percent	N	Percent	N	Percent	N	Percent	N
Key Trend Profiles								
Overall Rate Reduction and Break Toward Greater Improvement	10.60%	58	14.29%	44	11.49%	47	11.79%	149
Overall Rate Reduction and No Break or Break Toward Less Improvement	34.00%	186	33.44%	103	31.05%	127	32.91%	416
No Rate Change	46.44%	254	41.56%	128	45.23%	185	44.86%	567
Overall Rate Increase	8.96%	49	10.71%	33	12.22%	50	10.44%	132

Source: The Evaluation Contractor Analysis of HEN Monthly Report Data, November 2014 and HEN timeline data, fall 2014.
[a]The number of measures meeting 17.6 percent/10 percent goals excludes measures that also met the 40 percent/20 percent goals. This total number of measures differs from the ITS results presented in Chapter 4, where the number of measures meeting the 17.6 percent/10 percent goals includes the measures that also met 40 percent/20 percent goals.

Next, the results of the interrupted time series analyses—specifically the percentage of measures with a significant break toward greater improvement, a break toward less improvement, and no significant break—were compared across HEN activity levels. HENs with medium and high activity levels exhibited larger percentages of measures with structural breaks toward greater improvement (26.30 and 25.43 percent, respectively) than HENs with low activity levels (17.92 percent). Additionally, HENs with low levels of activity were more likely than other HENs to have measures that exhibited no significant structural breaks during the campaign (67.64 percent). However, in another counterintuitive finding, HENs with high activity rates also demonstrated larger percentages of measures with structural breaks toward less improvement (18.58 percent).

Finally, the percentages of measures falling into trend profiles formed by combining the overall change in rate from the beginning to end of the series, with the presence and direction of significant structural breaks, were compared across HEN activity levels. HENs with medium and high levels of activity had slightly higher percentages of measures with the better trend profile (i.e., percentage of measures that showed improvement and a break toward more improvement during the campaign.) However, again, these higher activity HENs also exhibited larger percentages of measures worsening. Across the three groupings of HENs, there were similar percentages of measures demonstrating overall improvements without structural breaks toward greater improvement and larger percentages of measures with no significant change overall.

The results presented in Table 5-9 are consistent with results observed for the individual AEAs. The detailed tables showing results for each AEA are located in Appendix E. Overall, the results suggest that HENs engaged in higher levels of activity were likely to see slightly larger percentages of measures improve and sometimes achieved greater reductions in patient harms. In contrast, however, these same high activity HENs were also more likely to exhibit larger percentages of measures worsening over time. Low activity HENs did not achieve as many improvements as their more active counterparts, but also did not have as many worsening outcomes. These results support the notion that engaging in more quality improvement activities is likely to lead to more successes, but may also produce more failures. Those that engage in fewer quality improvement initiatives may not see as many improvements, but are also unlikely to see any worsening rates.

Table 5-10 presents a summary of the ITS results by AEA. The table presents only those results that are consistent with the expected dose-response relationship. For example, if the percentage of hospitals meeting the 17.6 percent/10 percent improvement goals for a particular AEA was lowest for low activity HENs, and highest for high activity HENs, then that AEA is listed in the table. Detailed results for each AEA are located in Appendix E. The Evaluation Contractor cautions that this method of assessment assumes that there is a linear relationship between the activity level and outcome performance. There are several instances in which the medium activity group of HENs exhibits performance similar to or worse than the low activity, or similar to or better than the high activity HENs. This pattern of results may indicate that there is a non-linear relationship between HEN activities and outcome measure performance, in which outcomes improve after being exposed to at least some minimum threshold level of activity. At this time, however, there is no theoretical justification for selecting a specific threshold of activity to examine, nor is there sufficient data to control for the differences in the quality and effectiveness of activities across different HEN initiatives.

Table 5-10—Summary of ITS Results Consistent with Dose-Response Relationship Between HEN Activity Level and Performance Outcome Level, by AEA and Metric

Goal Performance	
Percentage of Measures that Met 17.6 Percent/10 Percent Goal	ADE
Percentage of Measures that Met 40 Percent/20 Percent Goal	OB-Other Pressure Ulcers Readmissions
Percentage of Measures that Did Not Meet Goal	Pressure Ulcers Readmissions
Measures Tested for Overall Change	
Average Change in Rate (Improving)	Readmissions VAP/VAE
Average Change in Rate (No Sig Change)	OB-Other Pressure Ulcers Readmissions VTE
Average Change in Rate (Worsening)	None
Changes in Measure Trends (ITS)	
Percentage of Measures with Break Toward Greater Improvement	CAUTI OB-Other Readmissions VTE
Percentage of Measures with No Significant Break	ADE CAUTI OB-Other VAP/VAE
Percentage of Measures with Break Toward Less Improvement	Readmissions
Key Trend Profiles	
Overall Rate Reduction and Break Toward Greater Improvement	None
Overall Rate Reduction and no Break or Break Toward Less Improvement	None

Table 5-10—Summary of ITS Results Consistent with Dose-Response Relationship Between HEN Activity Level and Performance Outcome Level, by AEA and Metric	
No Rate Change	ADE CAUTI OB-Other
Overall Rate Increase	CLABSI

Source: The Evaluation Contractor analysis of HEN-level data from HEN monthly reports, November 2014, CMS/Medicare, National Healthcare Safety Network (NHSN), and NDNQI data.

The implication of these findings is that high activity HENs may have been more willing to try new and different techniques for reducing patient harms, resulting in both greater rewards and more frequent failure. Furthermore, the finding that high activity HENs exhibited more measures worsening may be the result of greater vigilance and detection developed by the HENs in their pursuit of reducing patient harms. An equally likely truth related to these findings is that the HEN-submitted data on activities performed throughout the PfP campaign may not be particularly reliable, such that results are frequently in unexpected directions and appear to be unstable with respect to the expected hypotheses. The data available at present are not capable of distinguishing between these potentially different factors driving the results.

Impact Analyses of HEN Alignment within HEN Subgroups (Bayesian and Medicare Patient Safety Monitoring System [MPSMS])

The Evaluation Contractor examined the impact of HEN alignment on four outcomes (VTE, pressure ulcer, CRBSI, and readmissions) within groups of HEN-aligned hospitals based on the characteristics of the individual HEN with which they are aligned: intensity (high/low), collaboration (yes/no), and HEN type (hospital association/system/other) applying a Bayesian analytical model. Results for the entire post-intervention period (2012-2013) are consistent across HEN-groups. There are a few differences among the different combinations in the case of pressure ulcers and CRBSIs, although the differences are minor and not consistent across specific HEN types. For example, hospitals aligned with low-intensity collaborative and low-intensity non-collaborative hospital association HENs were more likely to experience a 25 percent or higher reduction (relative to nonaligned hospitals) in pressure ulcers, but that was not true for hospitals aligned with the other two subgroups that provided low-intensity support (hospital association HENs using hospital collaboratives and HENs not part of a hospital association or health system and not using collaboratives). In the case of CRBSI rates, hospitals aligned with high-intensity collaborative hospital association HENs and low-intensity collaborative HENs had a higher likelihood of a 10 percent or greater reduction than non-aligned hospitals, but that was not true for other HENs using hospital collaboratives.

The Evaluation Contractor also performed HEN subgroup impact analyses of five outcomes (any adverse event, any adverse drug event [ADE], any general adverse event, any post procedural adverse event, and any hospital acquired infection) using the MPSMS data. In this case, the difference-in-differences analyses included two categories of HENs: HEN type (complex, hospital association, system, and other) and HEN size (fewer than 50 hospitals, 50 to 99 hospitals, 100 to 400 hospitals, and more than 400 hospitals). The overall findings are that, among the various subgroups analyzed, there was no evidence that a particular type of HEN had consistently better outcomes. Only one of the 40 difference-in-differences estimates (five outcomes and eight different subgroups) was statistically significant and it was positive, suggesting that the decline in any general adverse events in non–HEN–aligned small hospitals (fewer than 50 beds) was actually larger than those in HEN–aligned hospitals.

More detail about these analyses were shown in Chapter 4.

Impact Analyses of HEN Alignment within Hospital Subgroups

Impact Analyses of HEN Alignment within Hospital Subgroups

The Evaluation Contractor also applied a Bayesian analytical model to test for differential impacts on VTE, pressure ulcer, CRBSI, and 30-day all-cause hospital readmissions among subgroups based on 13 combinations of three hospital characteristics: critical access hospital (CAH) status; ownership (private, federal or non-federal government, and non-profit); and bed size (fewer than 100 beds, 100 to 199 beds, 200 to 399 beds, and 400 or more beds). The Evaluation Contractor hypothesized that smaller hospitals or CAHs may benefit more from aligning with a HEN. However, no evidence indicated that hospital characteristics were associated with higher program impacts. While in some cases, HEN-aligned CAHs and small private hospitals (fewer than 100 beds) had similar impacts to other hospital types, such as with VTEs, in examining pressure ulcers the Evaluation Contractor found that HEN-aligned CAHs were the least likely to reduce outcomes.

Using the MPSMS data, the Evaluation Contractor performed subgroup analyses using an array of hospital characteristics—ownership type (private, non-profit, government), hospital size (non-CAH hospital with more than 400 beds, non-CAH hospital with 200 to 399 beds, non-CAH hospital with 100 to 199 beds, and hospital with fewer than 100 beds), teaching status, region (northeast, Midwest, south, and west), rural or urban, and percentage of patients who are covered by Medicaid (over 25 percent). Among the 85 different comparisons of impacts within subgroups (that is, PfP impacts within, for example, hospitals with 200 to 399 beds), there were only three estimates that were statistically significant at the 5 percent level. In all three cases, the difference-in-differences estimates were positive, meaning that the decline in any post-procedural adverse event was greater in the non-aligned teaching hospitals and hospitals in the northeast and south regions. The lack of plausible explanations for these three results, coupled with the fact that one would expect more than three statistical tests to be falsely significant among the tests conducted due to chance alone, suggests that these three estimates do not reflect any true pattern of hospitals that benefitted more from HEN alignment.

Variation in Outcome Trend Results, by HEN

Interrupted Time Series Results by HEN

The PfP Campaign involved 30 HENs and HEN cohorts working to spread best practices and innovative strategies in reducing patient harms across the 11 AEAs. Rather than assessing the change in a national rate, the HEN-level data provides the detail to understand which HENs achieved success, and in which topic areas. Additionally, the increased granularity of the HEN-level measures allows assessment of the variability in results across HENs. This section of the report presents an assessment of each HEN's data, to better understand which patient harms were reduced in more HENs, by greater amounts, and whether or not the aggregated trends across measures provide evidence that would be consistent with quality improvement changes occurring during the time frame of the PfP campaign.

In this section, the Evaluation Contractor reports results from the same group of analyses reported in Chapter 4 for the AEAs, but stratified by HEN. Again, outcome measures were evaluated for quality in terms of being widely and consistently reported among HEN hospitals. The extent of improvement in terms of whether or not measures achieved goal and comparing the percentage of measures with significant improvements to those that significantly worsened were examined next. The Evaluation Contractor assessed outcomes with sufficient observations using interrupted time series analysis to detect the presence and direction of any structural breaks in the measure series. Finally, the timing and direction of structural breaks was examined to better understand the contexts of the quality improvements in patient safety that were realized during the course of the PfP campaign.

Results for All HENs

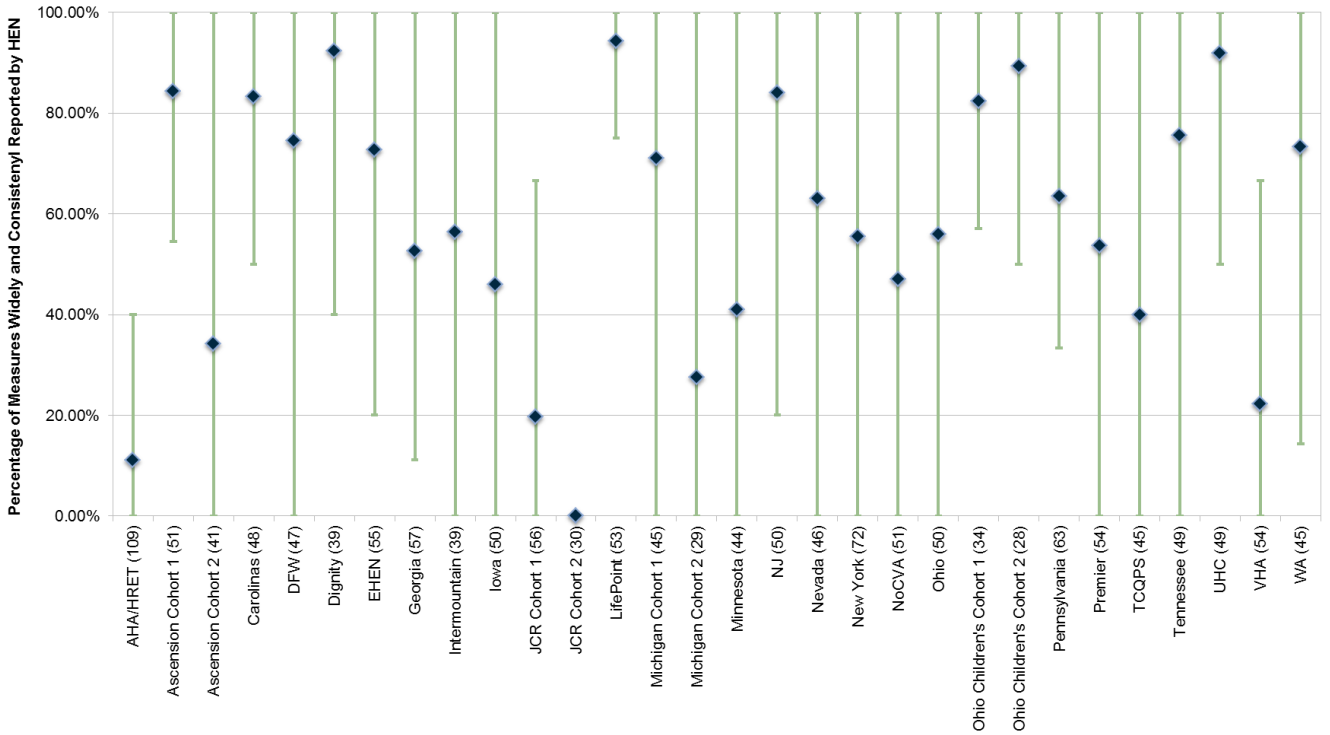
The HENs in the PfP campaign submitted 1,483 outcome measures. In nearly all cases, the HENs submitted data for all 11 AEAs.[5-28] The total number of measures reported ranged from 28 from Ohio Children's Cohort 2 to 109 from AHA/HRET, with a median of 49 measures per HEN. Results are scaled to percentages for the following analyses to permit comparison. The analyses reported previously were not repeated; the results were simply viewed from the point of view of HENs rather than AEAs.

Widely and Consistently Reported Measures

Figure 5-5 compares the percentage of HEN measures that were reported widely and consistently by HEN hospitals. Widely reported measures have at least 60 percent of aligned hospitals in the HEN reporting for at least 50 percent of the observations in the series. Consistently reported measures have hospital counts that are within, at most, 15 percent of the maximum hospital count in the series. As described previously, overall 56.91 percent of measures (n = 844) were widely and consistently reported. When these were distributed among the HENs, results ranged from a low of zero percent of measures widely reported for JCR Cohort 2 to a high of 94 percent of measures widely reported for LifePoint. The averages for the HENs varied widely, although most were above 40 percent.

[5-28] Ohio Children's Cohorts 1 and 2 did not report any measures for OB-EED; UHC's single measure was excluded due to mathematical error.

Figure 5-5—Percentage of Measures Widely and Consistently Reported, by HEN



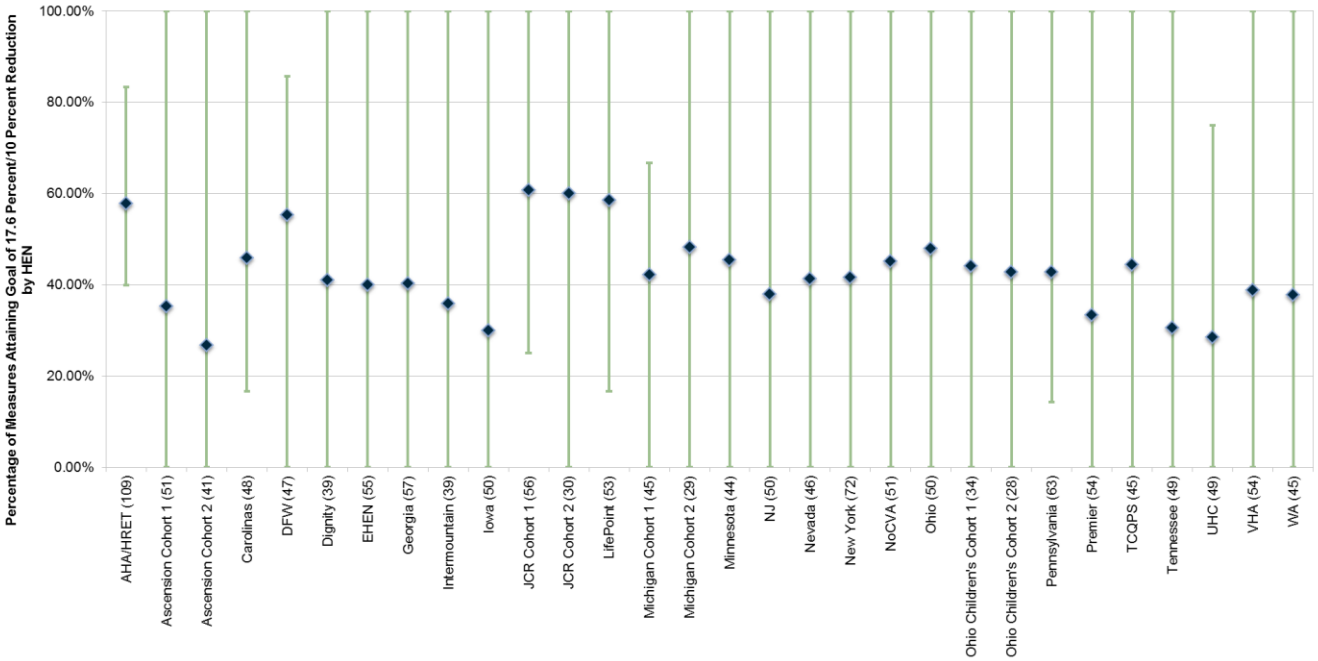
Source: Evaluation Contractor analysis of HEN-level data submitted by HENs, CMS/Medicare, NHSN, and NDNQI.

This figure also shows the ranges of results for each HEN across AEAs. Most HENs exhibited very wide ranges of results, indicating percentages of widely reported measures at or near zero for some AEAs, and at or near 100 percent for others. LifePoint had the narrowest range of variation, from 75 percent for at least one AEA to 100 percent for another. The next smallest range was achieved by a small group of HENs that had differences of approximately 45 to 50 percentage points between their worst showing and their best, which was 100 percent. These were Ascension Cohort 1, Carolinas, Ohio Children's Cohorts 1 and 2, and UHC. AHA/HRET also exhibited a relatively narrow range, but at the bottom of the scale, going from 0 percent to 40 percent of measures being widely reported. The 5 HENS with the highest average percentage of widely and consistently reported measures were LifePoint, Dignity, UHC, Ohio Children's Cohort 2, and Ascension Cohort 1.

Achieving Goal

Across the HENs and cohorts, the average percentage of measures meeting the 17.6 percent/10 percent reduction goal ranged from a low of just under 30 percent for Ascension Cohort 2 and UHC to a high of approximately 60 percent for both JCR cohorts. Half of the HEN cohorts attained goal for between 40 and 50 percent of their measures, as plotted in Figure 5-6.

Figure 5-6—Percentage of Measures Attaining Goal of 17.6 Percent/10 Percent Reduction, by HEN



Source: Evaluation Contractor analysis of HEN-level data submitted by HENs, CMS/Medicare, NHSN, and NDNQI.

The range of results for this analysis was exceptionally wide; with 22 of 30 HENs and cohorts encompassing everything from 0 percent to 100 percent of measures meeting goal, depending on the AEA. The only HEN with a substantially narrower range of variation was AHA/HRET, which reached goal for at least 40 percent of measures in at least 1 AEA, but for no more than 83 percent of its measures in its best-performing AEA.

Only 2 HENs exceeded an average level of performance of 50 percent: DFW, with an average of 55.32 percent, and JCR Cohort 2, with 60 percent of measures reaching goal. Conversely, only 3 HENs achieved an average of 30 percent or less of measures reaching goal for at least one AEA: Ascension Cohort 2, Iowa, and UHC.

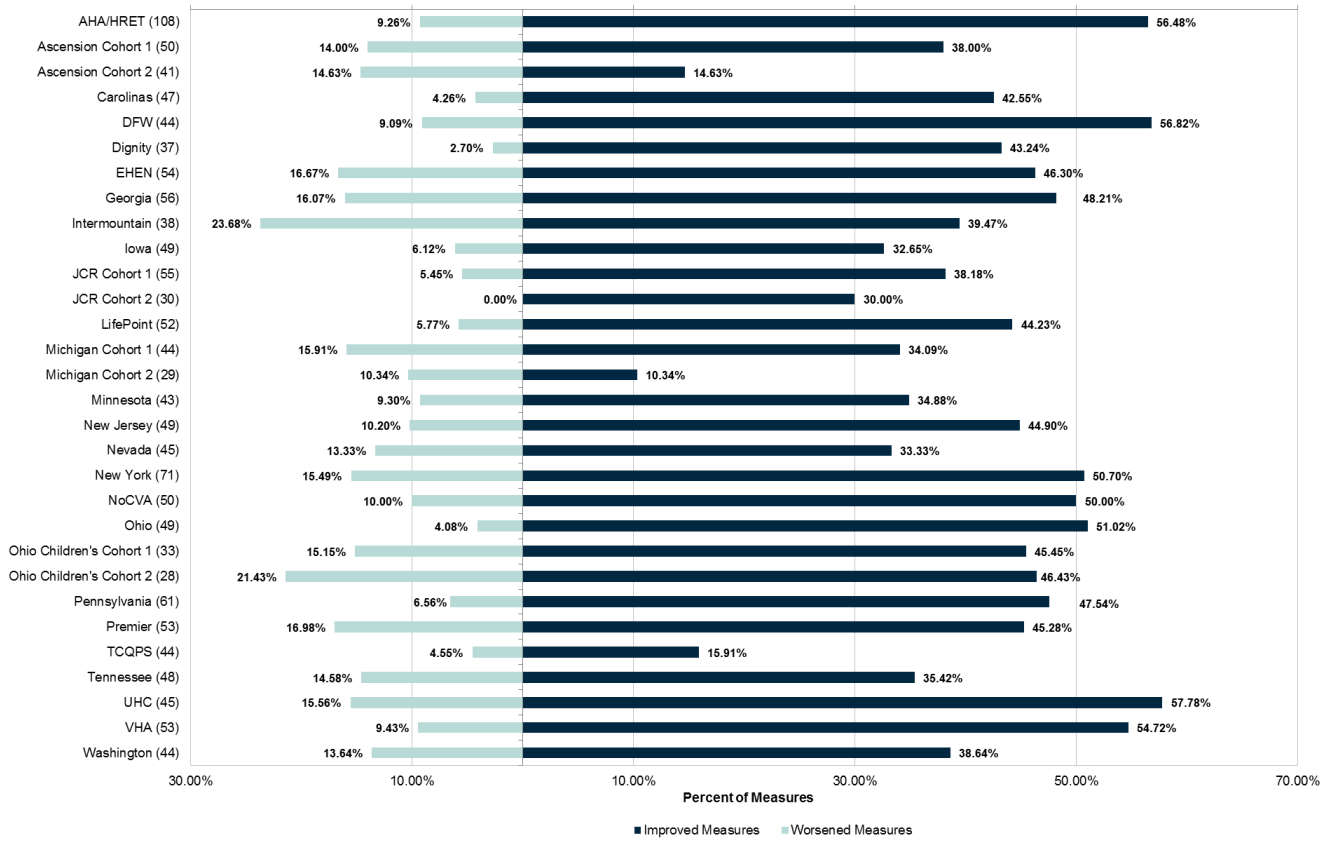
Statistically Significant Changes

Of the 1,483 outcome measures examined, 1,450 contained sufficient information to determine whether the measures exhibited a statistically significant change. These measures could be classified as either improving or worsening if the most recent 90 days of data were significantly different from the baseline, or not changing if the difference was not significant. Out of the 1,450 measures tested, 42.48 percent (n = 616) exhibited significant improvement, while 11.03 percent (n = 160) exhibited significant worsening. The distribution of these results across HENs and cohorts is provided in Figure 5-7.

Most of the HENs and cohorts reported significant improvements for between 30 percent and 50 percent of their measures. Four were measurably higher (AHA/HRET, DFW, UHA, and VHA), with between 54.72 percent and 57.78 percent of measures showing significant improvement. Three HENs were measurably lower, with Ascension Cohort 2 exhibiting significant improvement for only 14.63 percent of its measures, Michigan Cohort 2 for 10.34 percent, and TCQPS for 15.91 percent.

The percentage of measures that worsened significantly is smaller than those that improved for nearly all of the HENs, ranging between 0 percent for JCR Cohort 2, to 23.68 percent of measures worsening for Intermountain, as shown in Figure 5-7. Most of the HENs' performances were only between 0 percent and 17 percent of measures worsening. Overall, the ratio of improving measures to worsening measures was 3.85. Every HEN and cohort reported as many or more measures improving significantly as worsening significantly. The smallest ratio of improved to worsening measures was 1.0 for both Ascension Cohort 2 and Michigan Cohort 2.

Figure 5-7—Percentage of Significantly Improving Versus Significantly Worsening Measures, by HEN



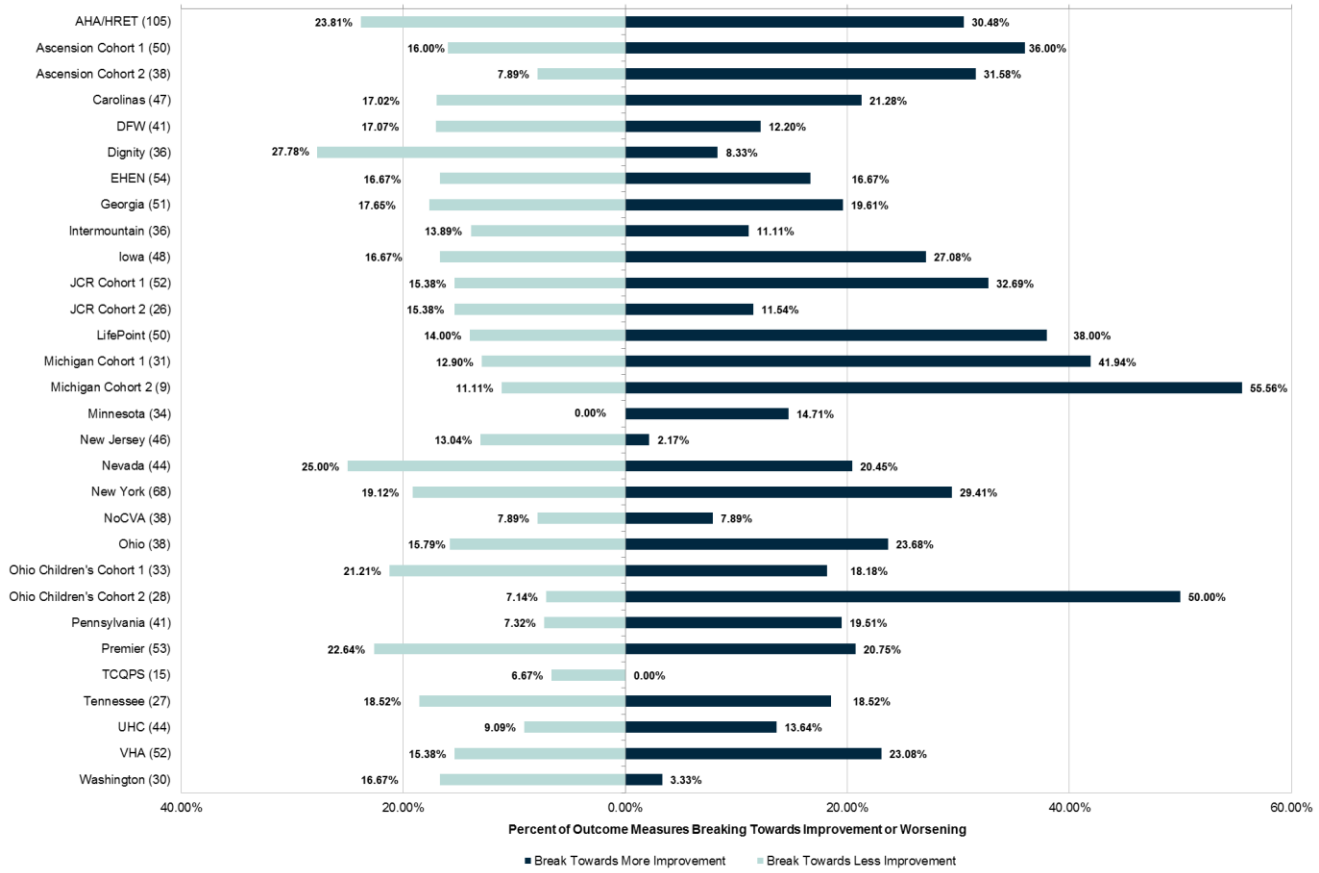
Source: Evaluation Contractor analysis of HEN-level data submitted by HENs, CMS/Medicare, NHSN, and NDNQI.

ITS Results

Across all HENs and cohorts, there were 1,265 measures with at least eight data points, the minimum threshold used in this report for ITS analysis, as discussed in the methodology in Appendix D. Among the interrupted time series models estimated, 22.37 percent (n = 283) of measures exhibited structural breaks associated with a change in trend toward greater improvement. Only 15.97 percent (n = 202) exhibited structural breaks associated with changes in trend toward less improvement. Figure 5-8 compares, on a HEN and cohort level, the percentage of measures that exhibited a significant break toward improvement (i.e., greater reduction in measure rates) to those exhibiting a significant break toward worsening (i.e., lesser reductions in measure rates).

The percentage of measures with breaks toward greater improvements tended to outweigh the percentages with breaks toward less improvement for two-thirds of the HENs, but 10 of the 30 had more breaks toward less improvement. However, only 6 HENs exhibited greater than 3 percentage points' difference between the two categories: Dignity, JCR Cohort 2, New Jersey, Nevada, TCQPS, and Washington. Figure 5-8 illustrates the percentage of measures with structural breaks toward greater improvement compared to those with breaks toward less improvement (or worsening).

Figure 5-8—Percentage of Outcome Measures Breaking Toward Improvement or Worsening, by HEN

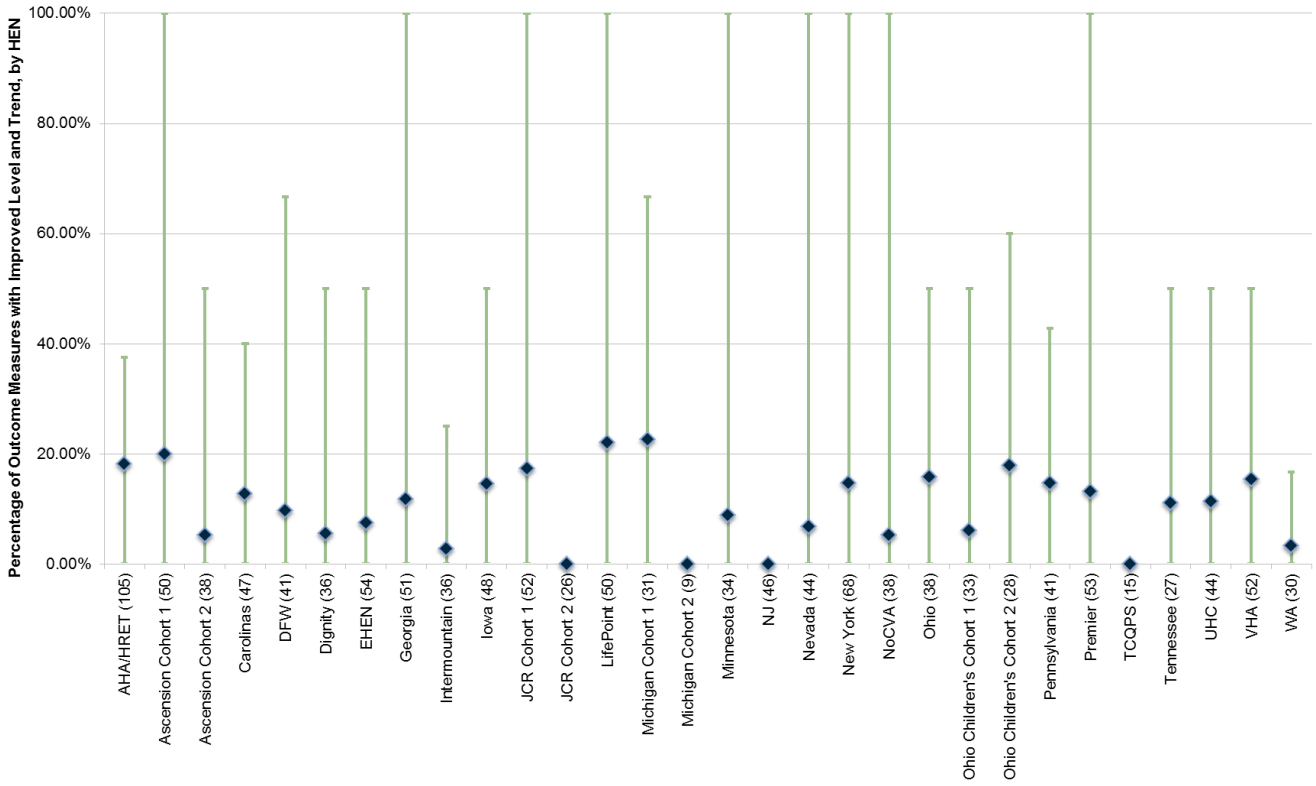


Source: Evaluation Contractor analysis of HEN-level data submitted by HENs, CMS/Medicare, NHSN, and NDNQI.

To focus more precisely on those measures that showed the best evidence for changes that were consistent with improvements in the processes driving patient harms during the PfP campaign, the next analysis sought to identify measures with both an overall improvement in rate from baseline and a structural break toward greater improvements during the PfP time frame. Of the 1,265 measures examined, 11.78 percent (n = 149) exhibited both of these characteristics. The relative percentages of measures for the 26 HENs and cohorts that met these criteria are provided in Figure 5-9.[5-29]

The averages by HEN are more tightly clustered across AEAs in this analysis than the others and tend to exhibit lower percentages of measures in this high category, ranging from 2.78 percent for Intermountain to 22.58 percent for Michigan Cohort 1. The 5 highest performing HENs by these criteria were Michigan Cohort 1, LifePoint, Ascension Cohort 1, AHA/HRET, and Ohio Children's Cohort 1.

Figure 5-9—Percentage of Outcome Measures with Improved Level and Trend, by HEN

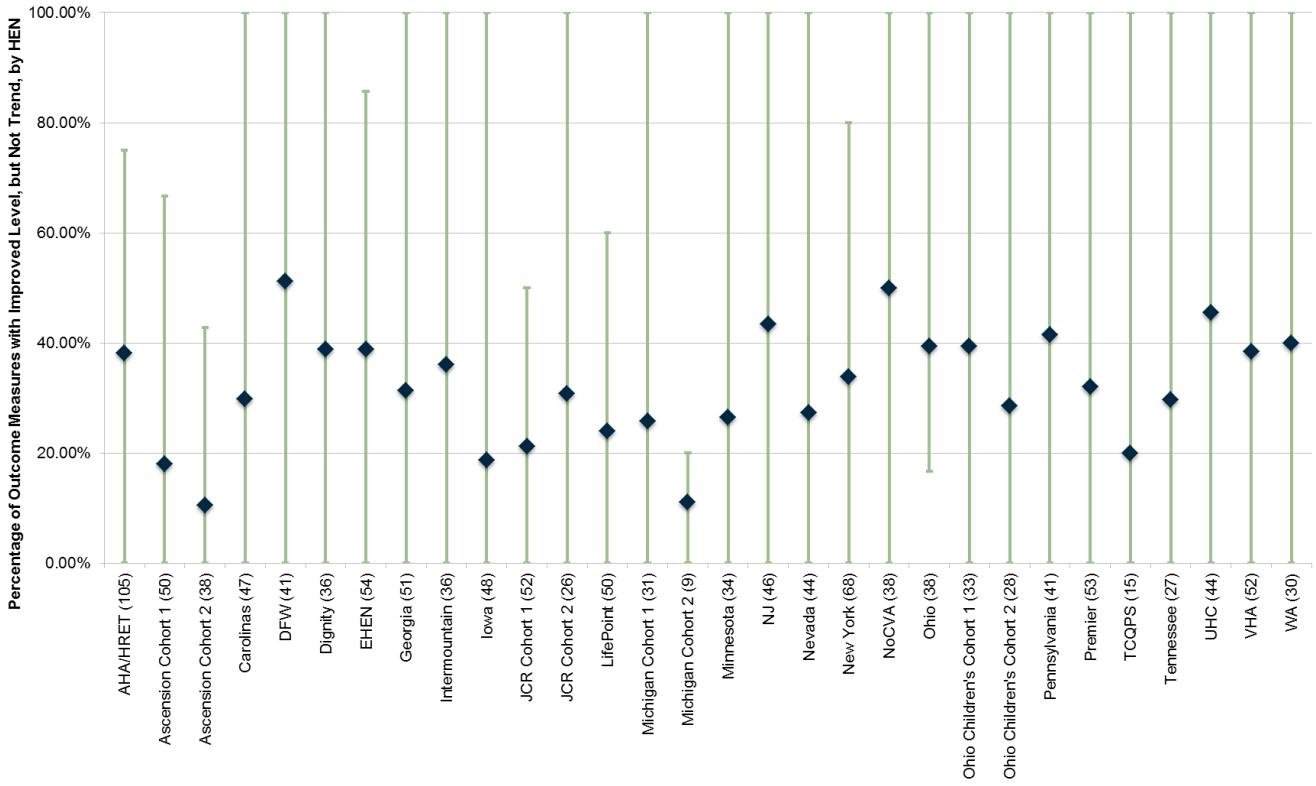


Source: Evaluation Contractor analysis of HEN-level data submitted by HENs, CMS/Medicare, NHSN, and NDNQI.

5-29 JCR Cohort 2, Michigan Cohort 2, New Jersey, and TCQPS did not have any measures in this group.

Finally, the Evaluation Contractor identified the percentage of measures in each HEN that improved significantly over the course of the PfP campaign, with either no significant break in trend or with a significant break toward less improvement. This provides a secondary way of examining the HENs that improved, as it includes improving measures that do not fit the ideal pattern of exhibiting a break toward more improvement. This data is provided in Figure 5-10.

Figure 5-10—Percentage of Outcome Measures with Improved Level, but Not Trend, by HEN



Source: Evaluation Contractor analysis of HEN-level data submitted by HENs, CMS/Medicare, NHSN, and NDNQI.

By this measure, all of the HENs exhibited measures improving in level, with most of the results clustered between 20 and 40 percent of measures showing improvements. Five HENs performed better: DFW (51.22 percent), NoCVA (50 percent), UHC (45.45 percent), New Jersey (43.48 percent), and Pennsylvania (41.46 percent). At the other end of the scale, Ascension Cohorts 1 and 2 both exhibited less than 20 percent (10.53 and 18.00 percent respectively), as did Michigan Cohort 2 (11.11 percent) and Iowa (18.75 percent).

Again, most of the HENs exhibited ranges in results encompassing results from zero to 100 percent of measures improving, depending on the AEA. Those HENs with narrower ranges for this analysis tended to be on the lower end of the continuum of results, starting at zero and improving only up to 20 percent for Michigan Cohort 2, the cohort that exhibited the narrowest range of performance results.

Taken together, certain HENs tend to rise to the top in each of these analyses. The combined results are displayed in Table 5-11 with HENs listed in alphabetical order.

Table 5-11—Top Five HENs According to Each Performance Metric

Percentage of Measures Widely-Reported/ Consistent Measures	Percentage of Measures Meeting 17.6 Percent/10 Percent Goal	Ratio of Improving Measures to Worsening Measures	Percentage of Measures with Improvement in Both Level and Trend	Percentage of Measures with Improved Level But not Trend
Ascension Cohort 1	AHA/HRET	Carolinas	AHA/HRET	DFW
Dignity	DFW	Dignity	Ascension Cohort 1	New Jersey
LifePoint	JCR Cohort 1	JCR Cohort 2	LifePoint	NoCVA
Ohio Children's Cohort 2	JCR Cohort 2	LifePoint	Michigan Cohort 1	Pennsylvania
UHC	LifePoint	Ohio	Ohio Cohort 1	UHC

Source: Evaluation Contractor analysis of HEN-level data submitted by HENs.

Notable Performance Stories

The PfP campaign was designed with bold aims in mind for improvement on a national scale, aligning every possible public and private sector force to improve patient safety. CMS invested considerable resources to provide the assistance necessary to HENs and individual hospitals to achieve these aims, and the work was managed aggressively and continuously to reach the highest level of leadership engagement, commitment and performance possible. A variety of approaches were used by CMS to provide continuous feedback to the HENs, including a platform for networking and interaction with regular webinars (such as pacing events), a strong centralized web resource, and regular challenges to the HENs to increase their levels of commitment and performance.

The following section highlights notable examples of patient safety reductions achieved by some of the HENs participating in the PfP campaign. These stories focus on examples from the common measures that demonstrate the intensive work done by the HENs, and, to some extent, show how they interacted with the engines of change harnessed for this campaign.[5-30] The HEN detail was gathered from timeline data in which the HENs described the initiatives they sponsored during the PfP campaign.[5-31] Their descriptions were coded by the Evaluation Contractor into one of four broad categories: tools, education, coaching, or leadership. HENs were also asked to name any external partners with which they collaborated, and describe, among other things, whether they had worked with the partner on the specific AEA before the PfP campaign, and whether they found the partnership valuable. Both initiatives and partnerships could be identified as relating to a specific AEA, or to general topics such as safety across the board or culture of safety.

[5-30] Common measures are measures for which multiple HENs reported data using the same, or highly consistent, specifications. Common measures therefore allow a degree of comparison across the HENs with respect to performance on specific outcomes. A complete list of the common measures can be found in Appendix E.

[5-31] Further details regarding the HEN measure and timeline data are available in Appendix B.

The results are by no means a list of the best HENs, nor do they cover all eleven AEAs. Rather, this discussion considers evidence from SPC charts, ITS results, and the HEN timeline data to gain insight on the many approaches the HENs employed to improve patient safety between 2010 and 2014.[5-32]

Adverse Drug Events (ADEs)

ADE, with the widest range of different measures—212 separate measures across the course of the campaign, was the patient harm area that began with the least agreement on appropriate strategies. Only 3 HENs and 200 participating hospitals were reporting any data at all on this topic in early 2012. Thus one of the preliminary goals was building the population of hospitals that recognized ADE as an important issue for patient safety and encouraging the measurement and reporting of its occurrence.

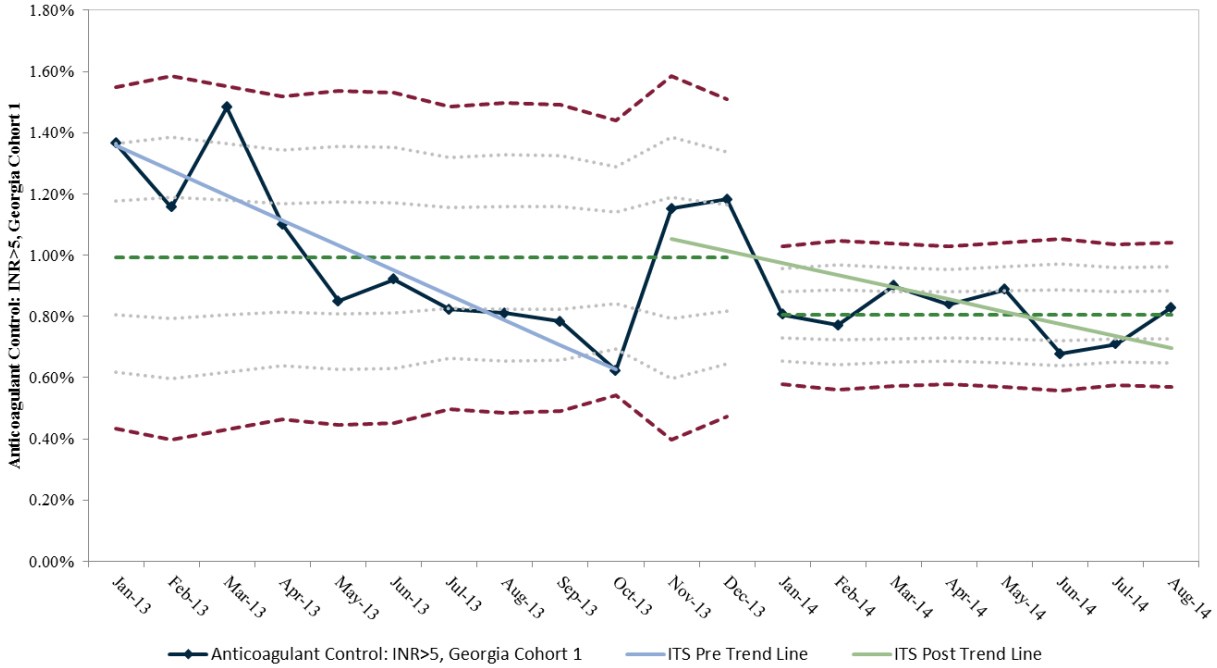
PfP included several concentrated drives on ADEs that reached all of the participating HENs. There was a pacing event in October 2012, followed by a campaign-wide push for data reporting. A series of workshops on the ADE measures were hosted by the National Content Developer (NCD) between January and March 2013. In December 2013 PfP leadership requested that all HENs focus on reporting data for three specific ADE areas: glycemic control, anticoagulants, and opioids. As the PfP campaign progressed, as many as 27 HENs and cohorts and 712 hospitals reported data for ADE measures in a single period.

Georgia's results illustrate how these PfP program elements converged, and exemplify the significant improvements in this AEA. Georgia was not among the early adopters of measuring ADE, not reporting measures until Q1 2013. These included readings of the international normalized ratio (INR) indicating possible clotting abnormalities for patients on warfarin (INR>5), and the incidence of blood glucose readings below the threshold level of 50 mg/dL (BG<50mg/dL) for patients prescribed insulin. The HEN started with relatively high baselines for the two measures—1.37 percent of readings from eligible patients exhibited INR>5 in January 2013; 0.71 percent of readings from eligible patients exhibited levels BG<50 mg/dL. By the end of the PfP campaign, Georgia had made significant reductions in both outcomes and achieved a 44.82 percent reduction in patients with INR>5, and a 39.08 percent reduction in readings with BG<50 mg/dL.

Figure 5-11 and Figure 5-12 show the general downward trends for the two ADE measures. For the first measure, patients with INR>5, the SPC results display evidence of improvement by a shift of the center line in January 2014. There was a decrease in the rate of INR>5 from January 2013 to August 2014, when it reached 0.83 percent. The ITS analysis of this measure identified a significant structural break in November 2013 toward a slower the rate of improvement. These two results are not contradictory; it is not unusual for gains to come quickly in the initial stages of improvement and slow later when the room for additional improvement is smaller.

[5-32] See Appendix D for details on SPC charts and ITS methodologies.

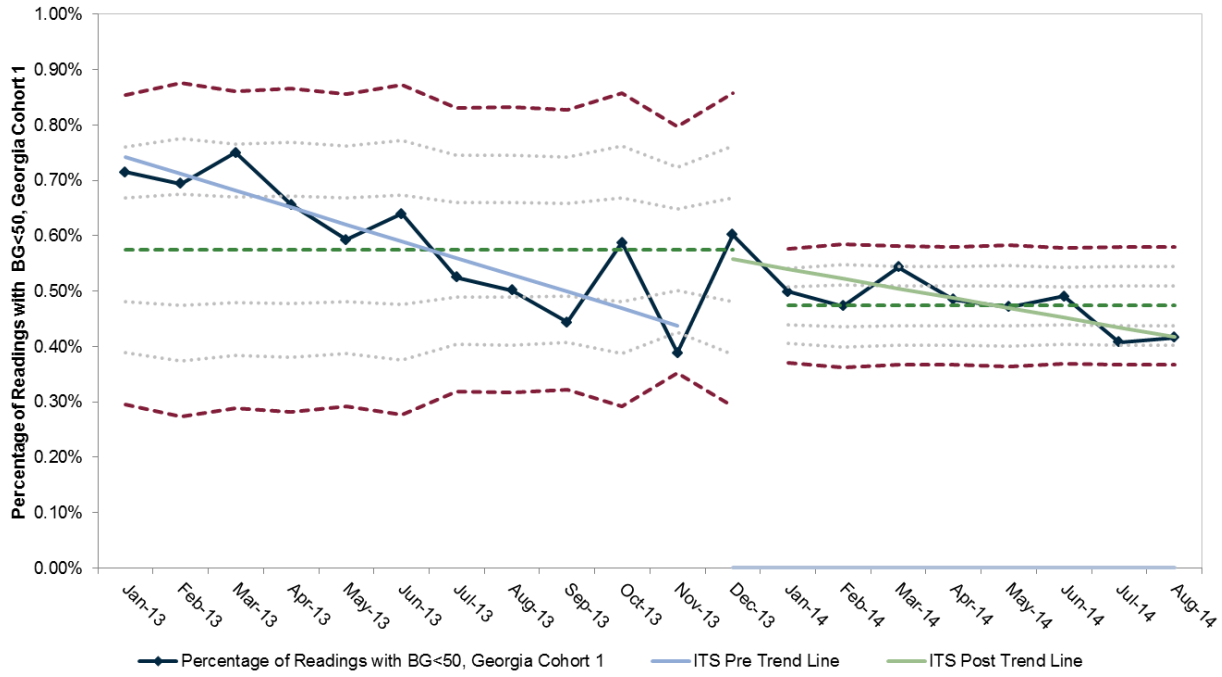
Figure 5-11—Anticoagulant Control INR>5, Georgia Cohort 1



Source: HEN Monthly Reports, November 2014 (n = 71 to 91 hospitals, depending on time period).
Note: Center line and control limits (U' chart) for first phase were calculated with data between January 2013 and December 2013. Center line and control limits (U' chart) for second phase were calculated with data between January 2014 and August 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Figure 5-12 shows very similar results for the second measure, BG<50 mg/dL. The SPC results parallel those for the INR measure, with a shift of the center line in January 2014. The results indicate an improvement in the rate of blood glucose readings below 50 mg/dL from 0.71 percent in January 2013 to 0.42 percent in August 2014, and the ITS analysis detected a significant structural break in December 2013 indicating some slowing in the rate of decrease.

Figure 5-12—Percentage of Readings with BG<50, Georgia Cohort 1

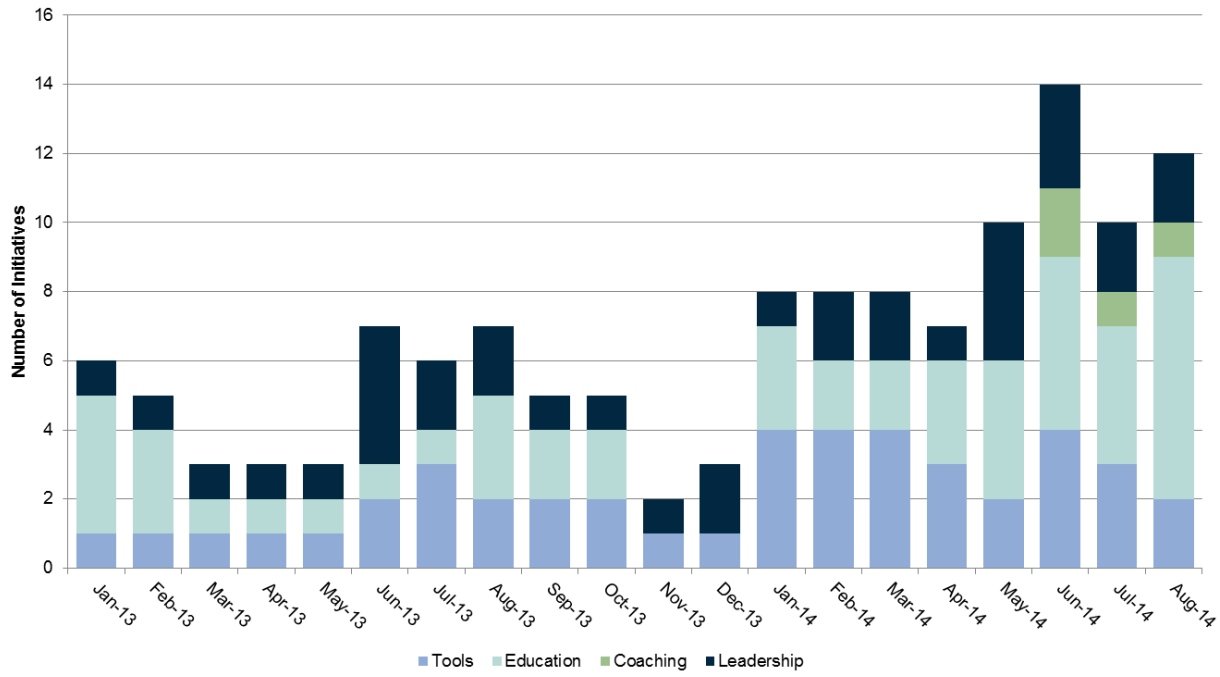


Source: HEN Monthly Reports, November 2014 (n = 56 to 84 hospitals, depending on time period).
Note: Center line and control limits (U' chart) for first phase were calculated with data between January 2013 and December 2013. Center line and control limits (U' chart) for second phase were calculated with data between January 2014 and August 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Georgia's timeline data showed extensive interaction and work at many levels with many different organizations. The HEN identified 22 separate partnerships as important to its work in this AEA. Partners included state and local public agencies, national private groups, professional associations, universities, and other HENs. The HEN reported entering into two partnerships specifically addressing this AEA. It joined a coalition or working group including the Iowa HEN beginning in March 2013 and continuing throughout the campaign. This partnership was a direct result of the work both HENs did in PfP, and Georgia reported that the collaboration added value to their approach. Georgia reported a similar (but more limited) partnership from March through May 2013 with the Texas QIO. Again, the two entities had no working relationship prior to PfP, but Georgia indicated that the QIO contributed value to their approach to ADE.

Figure 5-13 shows the HEN timeline data regarding the initiatives they employed to reduce ADE.

Figure 5-13—ADE Initiatives Timeline, Georgia



Source: HEN timeline data, fall 2014.

Employing a strategy that was common to many HENs, Georgia reported engaging in a variety of initiatives applicable to all of its AEAs at once, rather than focusing its initiatives on specific AEAs. The HEN's initiatives provided leadership training, Lean Six Sigma training, and training regarding the components and implementation of reliable process design. In general, the number of initiatives increased slowly and steadily, with a low of two initiatives in November 2013, and a high of 14 in June 2014, before tapering slightly as the PfP campaign ended. Thus, it appears that as the HEN engaged in more activities it saw improvements in patient safety over time. Once improvement started to slow, they increased their efforts, but with what may have been diminishing returns.

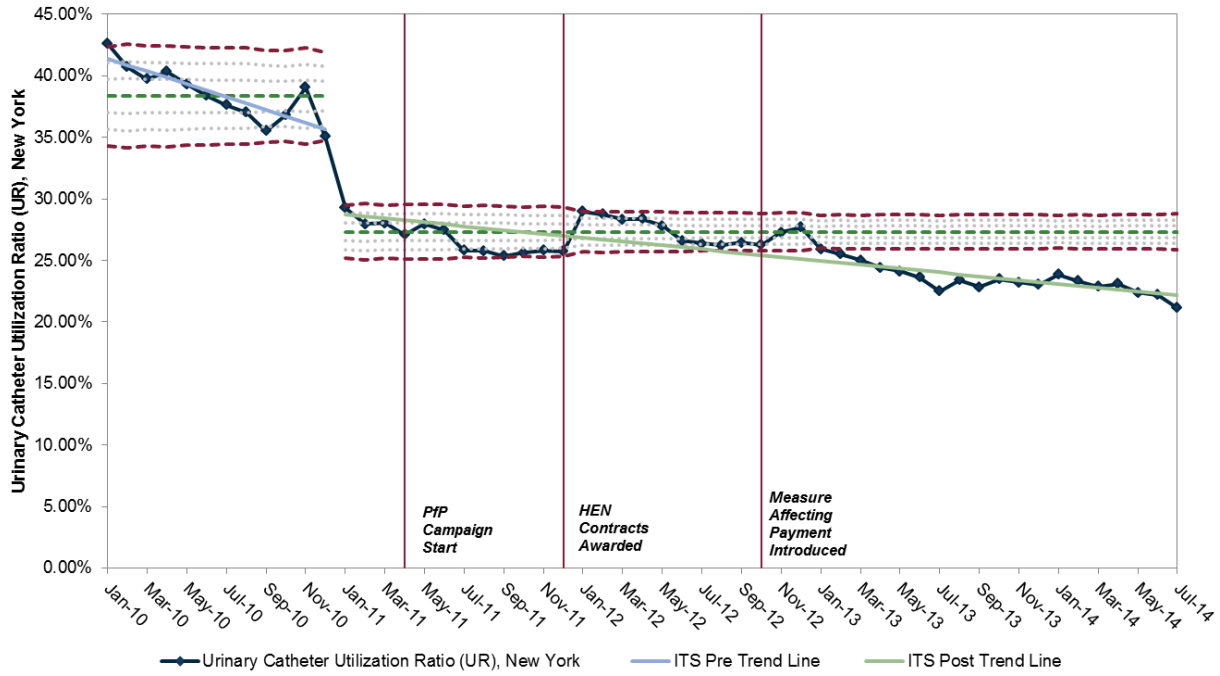
CAUTI and CLABSI

Over the course of the PfP campaign, the New York HEN showed exemplary performance in reducing the utilization of both urinary catheters and central venous lines. In addition, the HEN was successful in reducing CLABSI rates at the same time it was reducing central line utilization. Figure 5-14 illustrates what New York accomplished with its Urinary Catheter Utilization Ratio (UR), and is accompanied by Figure 5-15, which depicts the HEN's CAUTI initiatives.

New York had already made substantial improvements in its CAUTI UR prior to the PfP campaign, reducing the rate from the placement of urinary catheters in 42.61 percent of patients in January of 2010, to 28.05 percent of patients for the month preceding the announcement of the PfP campaign (March 2011), an improvement of 34.18 percent. The ITS results showed a structural break towards worsening improvement in January 2011. Thereafter, the pace of progress during PfP was relatively slow. Improvement appeared to stall in 2012, but picked up again in 2013 and 2014, as indicated by a long string of months that fell below the

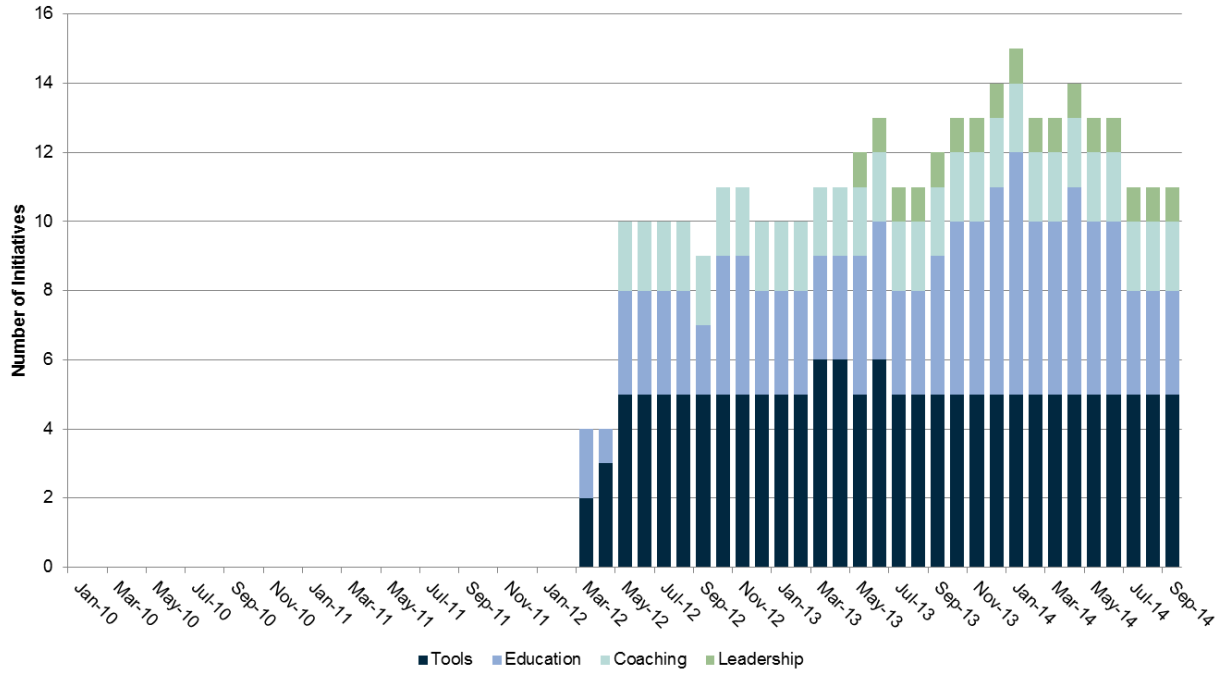
SPC chart lower control limits in Figure 5-14. By the end of September 2014, the UR had been reduced to 19.89 percent.

Figure 5-14—Urinary Catheter Utilization Ratio (UR), New York



Source: HEN Monthly Reports, November 2014 (n = 48 to 125 hospitals, depending on time period).
Note: Center line and control limits (U' chart) for first phase were calculated with data between January 2010 and December 2010. Center line and control limits (U' chart) for second phase were calculated with data between January 2011 and March 2012. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Figure 5-15—CAUTI Initiatives, New York



Source: HEN timeline data, fall 2014.
Note: No data available prior to January 2012. For January 2012 and later, no bar is shown when there is no targeted activity or general cross-cutting activity related to this specific measure or its AEA.

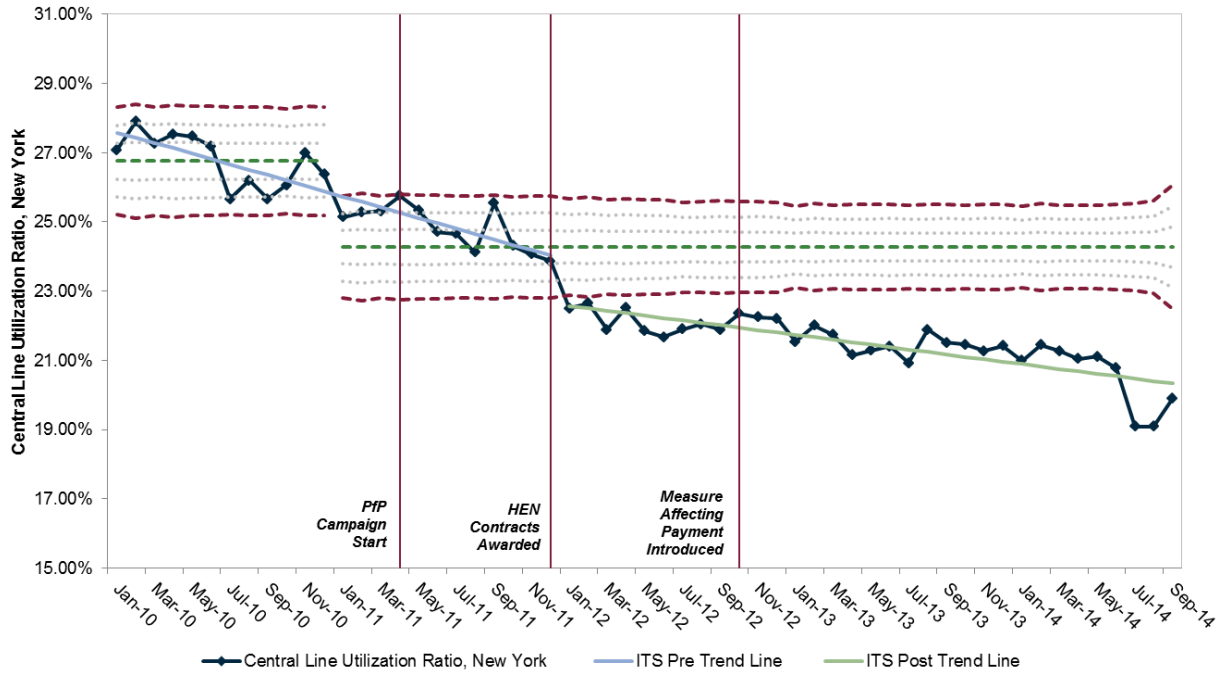
New York did not make these improvements without effort, as demonstrated by the initiative timeline data in Figure 5-15 above. Beginning in early 2012, the HEN engaged in multiple initiatives aimed at reducing CAUTI. In particular, they initiated a four-part strategy that included educational programs in mid-2012; a "Reverse the Trend ED/Hard Stop Pilots" initiative in mid-2013 that implemented pilot programs to reduce unnecessary urinary catheter use in the emergency department and other hospital units; a mix of coaching and "office hours" from late 2013 through early 2014 targeting five specific barriers to CAUTI reduction; and a webinar series in early to mid-2014 that also emphasized overcoming major barriers to CAUTI reduction. This set of targeted initiatives kicked off during the period of slowest progress in the UR and continued through the second period of sustained improvement towards the end of the campaign.

The HEN's reported partnerships included only one that was specifically geared toward CAUTI. The collaboration involved joint training with the state's QIO, IPRO, in on-site coaching for CAUTI in September 2013 in conjunction with the HEN's regional kick off meetings.

Turning to CLABSI, although New York's achievements in reducing central venous line utilization were not as dramatic as those in urinary catheter utilization, the HEN still made substantial and significant progress in this area. Figure 5-16 and Figure 5-17 illustrate the results of the SPC and ITS analyses for the utilization ratio and CLABSI rates per central line days, respectively. Prior to the start of PfP, the central line utilization ratio was reduced from 27.08 percent in January 2010 to 25.32 percent for March 2011, a 6.50 percent reduction, with the SPC chart showing a shift in the center line toward improvement in late 2010. The ITS results identified a structural break toward slower improvement in January 2012. During the PfP campaign, New York was successful in further reducing central line utilization to 19.91 percent by September 2014, a reduction of 26.47 percent relative to January 2010, and 21.36 percent relative to March 2011. While large

gains were made in the early months of the PfP campaign, progress slowed between mid-2012 and early 2014, although it may have begun improving again by June of that year.

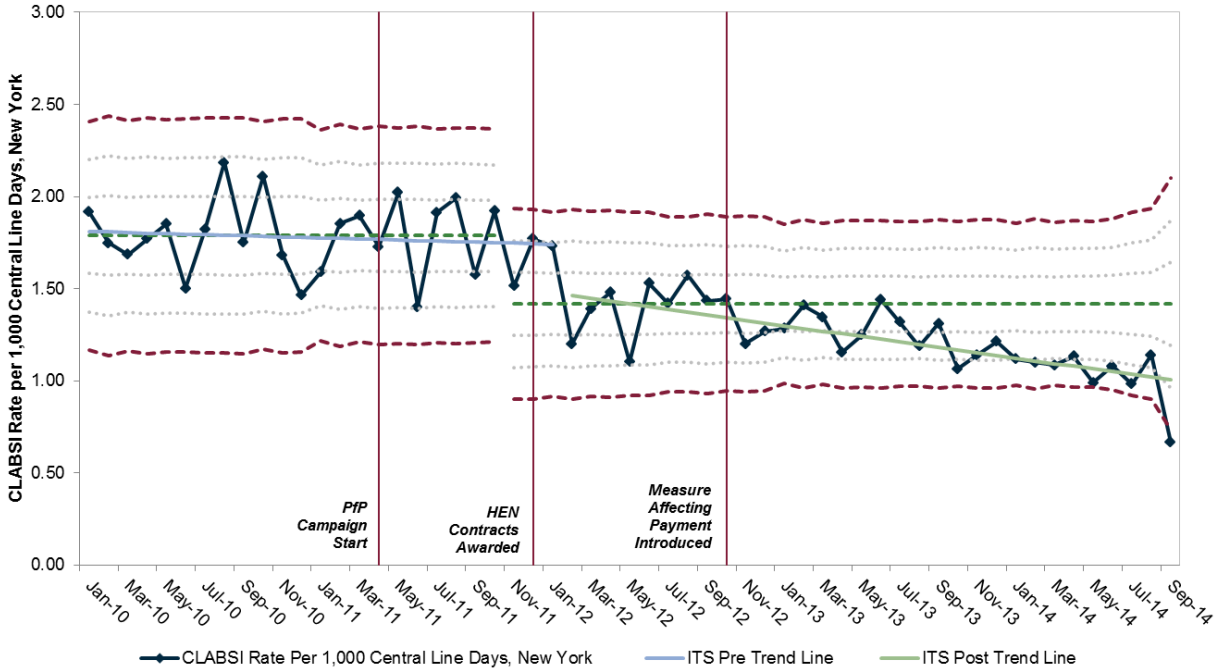
Figure 5-16—Central Line Utilization Ratio, New York



Source: HEN Monthly Reports, November 2014 (n = 54 to 123 hospitals, depending on time period).
Note: Center line and control limits (U' chart) for first phase were calculated with data between January 2010 and December 2010. Center line and control limits (U' chart) for second phase were calculated with data between January 2011 and March 2012. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

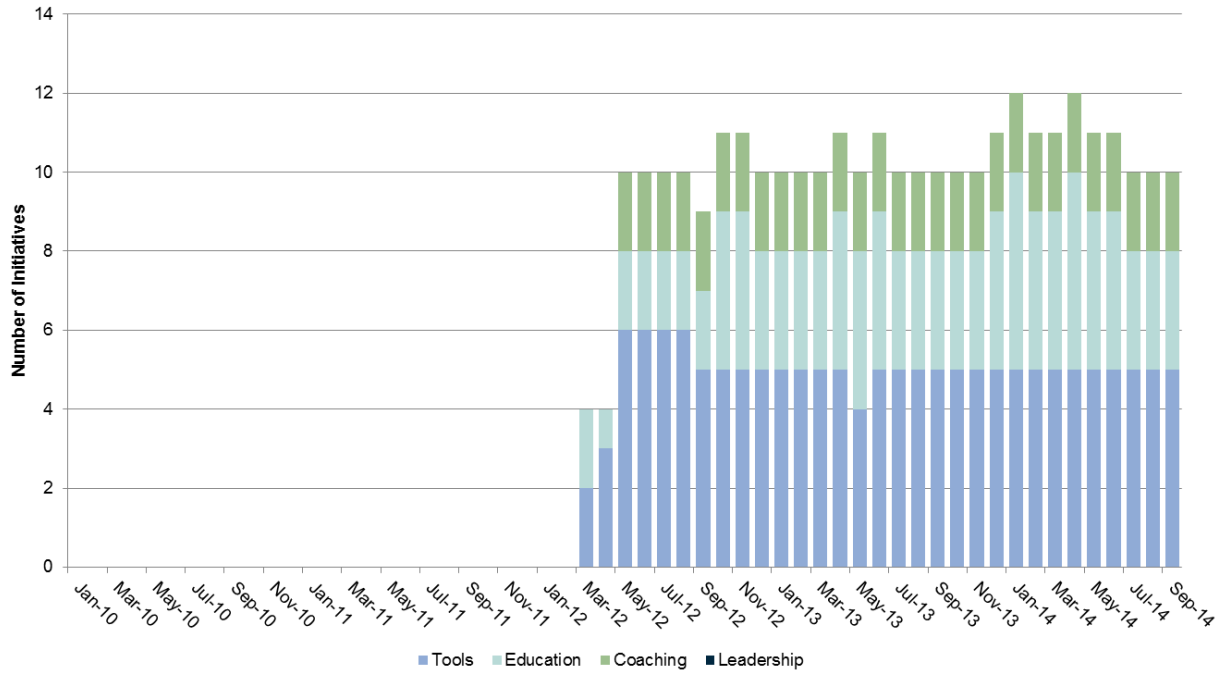
In contrast to its performance with CAUTI and CLABSI device utilization ratios, New York did not have a history of major reductions in its CLABSI infection rate prior to the advent of the PfP program. In fact, as seen below in Figure 5-17, the CLABSI rate of 1.92 per 1,000 central line days in October 2011 was identical to the rate of 1.92 per 1,000 in January 2010, indicating no real change for nearly two years. A substantial string of improvements began later in 2011, however, roughly at the same time the PfP program was launched. The CLABSI rate was reduced to 0.67 per 1,000 central line days by the final month for which data were reported (September 2014)—a 65.28 percent drop relative to January 2010. Although gains were relatively flat between January 2010 and January 2012, the ITS analysis detected a structural break towards greater improvement in February 2012, a finding consistent with additional quality improvement efforts being made among the HEN hospitals.

Figure 5-17—CLABSI Rate per 1,000 Central Line Days, New York



Source: HEN Monthly Reports, November 2014 (n = 54 to 123 hospitals, depending on time period).
Note: Center line and control limits (U' chart) for first phase were calculated with data between January 2010 and March 2011.Center line and control limits (U' chart) for second phase were calculated with data between November 2011 and January 2013.The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Figure 5-18—CLABSI Initiatives, New York



Source: HEN timeline data.
Note: No data available prior to January 2012. For January 2012 and later, no bar is shown when there is no targeted activity or general cross-cutting activity related to this specific measure or its AEA.

Taken together, New York's progress in reducing both device utilization and infection rates is remarkable. It seems likely that when hospitals reduce device utilization, they do so by eliminating unnecessary device use; the devices are usually least necessary in the patients that are most healthy. Thus, when device utilization shrinks, the remaining device users are sicker, and therefore more vulnerable to infection. This would tend to attenuate reductions in infection rates when utilization ratios come down, and this makes New York's success in bringing both outcomes down for CLABSI particularly notable.

New York's timeline data for cross-cutting and CLABSI-related initiatives and partnerships is displayed in Figure 5-18. Interestingly, New York's achievement in CLABSI occurred in the context of relatively few targeted efforts to reduce CLABSI, as reported by the HEN. These sessions focused on topics such as an insertion and maintenance bundle compliance and on non-critical care and long-term management of central lines. To be sure, these educational programs were not the sum total of New York's efforts in this area; for both CAUTI and CLABSI, the HEN provided aligned hospitals with coaching, guidance, feedback on their performance, and other resources and supports.

The only partnership referenced with regard to CLABSI was with the New York QIO, IPRO, already mentioned for its work on CAUTI. IPRO administered the Hospital Survey on Patient Safety Culture annually in 2012, 2013, and 2014.

Two changes in payment incentives may have influenced work in both CAUTI and CLABSI. In October 2011, the Inpatient Prospective Payment System (IPPS) payment incentive related to CAUTI went into effect, reducing payment for Medicare/Medicaid beneficiaries who sustained preventable CAUTI events. In October 2012, payment for hospital acquired CAUTI was ended altogether. If it is reasonable to expect payment penalties to drive rates downward, neither event appeared to statistically alter the existing trends for

improvement. CLABSI shows a similar pattern of results, although the payment incentives happened in October 2012 (IPPS payment incentive) and October 2013 (value-based purchasing program incorporated penalties for hospital-acquired CLABSI). Neither payment initiative had an apparent effect on the very gradual downward trend in CLABSI.

OB-EED

The prevention of OB-EEDs is perhaps the signature success of the all of the national efforts aimed at improving healthcare for mothers and infants from 2010 to 2014, and it became a cornerstone of the PfP campaign. Fully 21 HENs and HEN cohorts posted significant reductions in their rates of early deliveries that lacked medical indications.

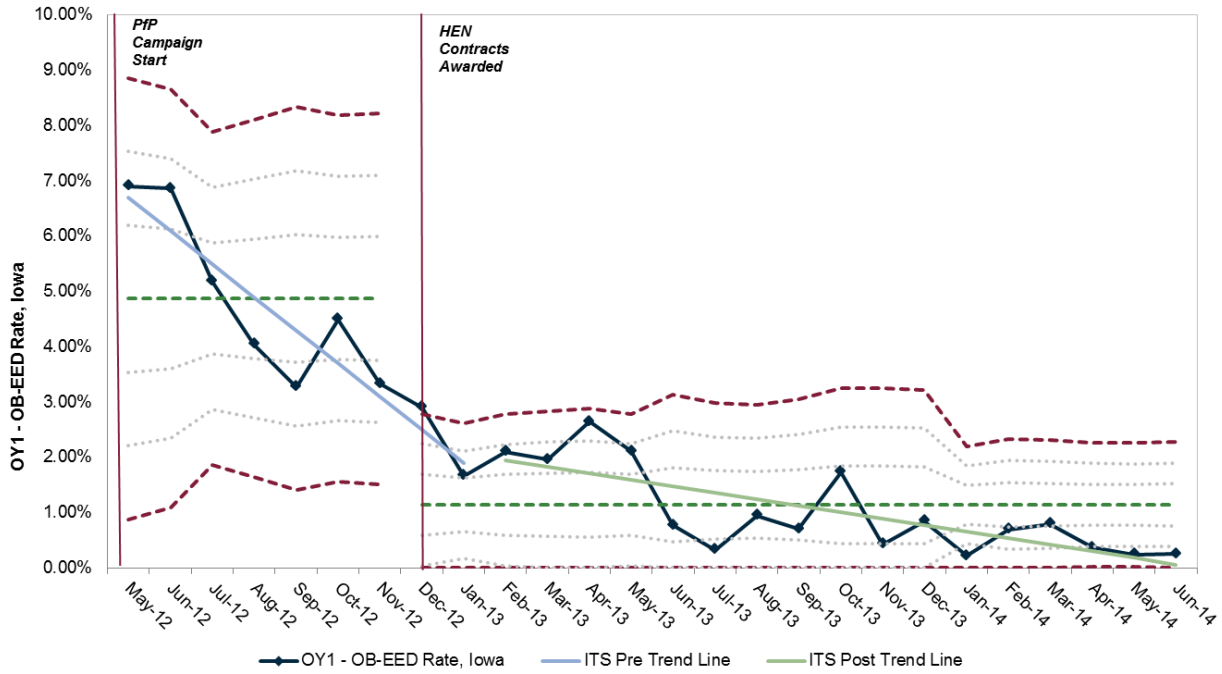
This widespread success may be related to two features unique to this AEA: an early understanding of the evidence-based best practice of limiting OB-EEDs, and widespread adoption of a hard stop policy. Unlike many of the other AEAs, there was wide agreement on a common measure early on, the Joint Commission's Perinatal Care (PC)-01. That measure captured the rate of non-medically indicated deliveries when the gestational age was greater than or equal to 37 weeks, and less than 39 weeks. The broad national consensus on a simple, bright line approach—the "hard stop" approach—may have been easier to implement than other more complicated improvements in other areas of patient harm.

The Iowa HEN's experience and OB-EED results are typical of the exceptional performance seen in this AEA. The HEN shared its progress with the entire PfP community in October 2013, when it presented evidence of the spread of the hard stop policy through its hospitals.[5-33] The percentage of HEN-participating hospitals employing the hard stop rose from 40 percent of 84 hospitals in June 2012, to 100 percent of that number by August 2013. Over approximately the same period, the HEN reported a decline in the percentage of OB-EEDs from a baseline average of just 6.91 percent of early deliveries in May 2012, to a rate of 0.78 percent in June 2013. The rates continued to improve generally through the campaign's end, achieving a rate of 0.25 percent by June 2014, a 95.31 percent improvement over the time period of the PfP campaign.

The SPC chart in Figure 5-19 shows evidence of an improvement in rate both in terms of a significant downward shift of the center line in December 2012, and in the fact that four of the last six observations fell between the 2-sigma line and the lower control limit. The ITS results evidence a significant structural break in February 2013 followed by a slower rate of improvement consistent with the measure approaching a floor in terms of eradicating early elective deliveries.

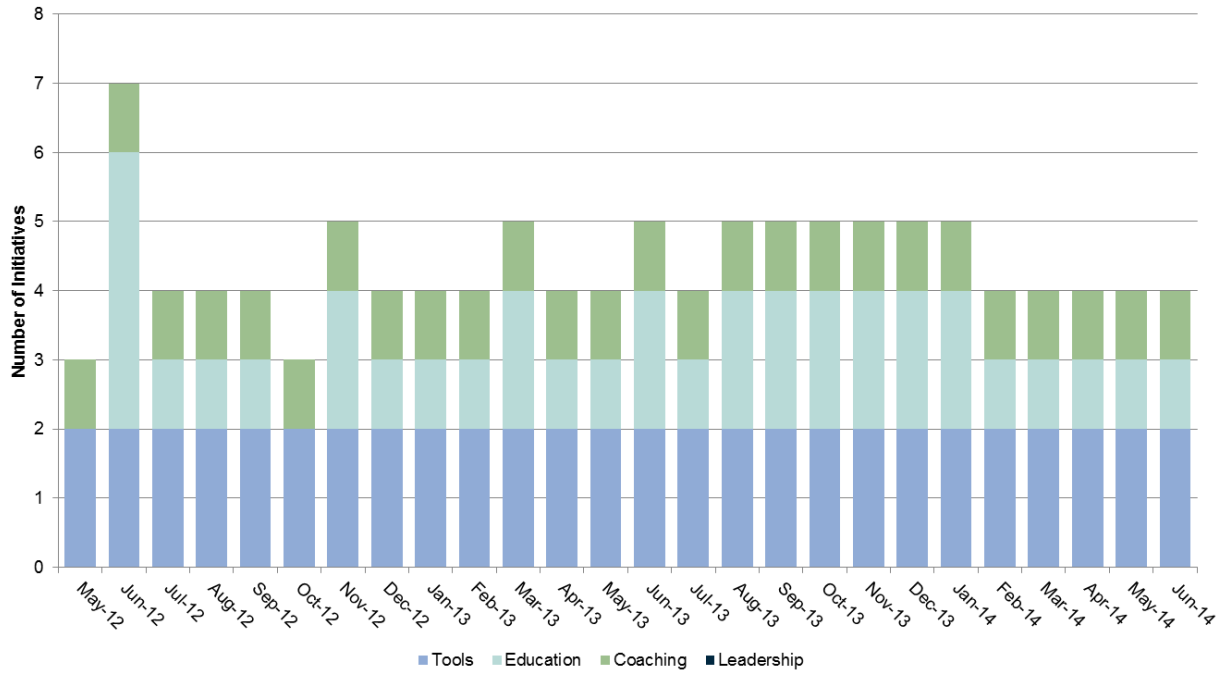
5-33 A video recording of this pacing event is available at: http://link.videoplatform.limelight.com/media/?mediaId=4d5f82e02d75407b97470ef68620717d&width=640 &height=480&playerForm=87808a650c524ab08129c8135ce5a8ec. Accessed on September 24, 2015.

Figure 5-19— OY1–OB-EED Rate, Iowa



Source: HEN Monthly Report, November 2014 (n = 50 to 62 hospitals, depending on the period).
Note: Control limits and center line (U chart) for first phase calculated with data from May 2012 to November 2012. Control limits and center line (U chart) for second phase calculated with data from December 2012 to June 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two sigma limits.

Figure 5-20—OB-EED Initiatives, Iowa



Source: HEN timeline data, fall 2014.

The HEN's work provides a good snapshot of the resources that were made available to the HENs participating in the PfP campaign. The HEN named 16 partners in this area, all members of a statewide OB Taskforce active at some level for the entire period from November 2012 through June 2014. The HEN's partners included a range of state and local public and private partners (e.g. the Iowa State Health Department, the Iowa Nurses Association, and the Iowa Academy of Family Physicians), as well as national private partners (e.g. March of Dimes, American Board of Internal Medicine [ABIM] Choosing Wisely, American Congress of Obstetricians and Gynecologists [ACOG]). Iowa identified its work with the State Health Department and ABIM as critical to their achievements.

As illustrated in Figure 5-20 above, Iowa also engaged in numerous initiatives. The HEN began reporting data and working on initiatives for OB-EED in May 2012. From that point through June 2014, 4 or 5 initiatives were active in almost all months. Initiatives related to tools, such as web pages with resources, and regularly updated web-based quality reports, predominated because they tended to extend over the length of PfP. However, the HEN also employed a significant mix of educational and coaching activities. Among these, the HEN made "improvement advisors" available for on-site coaching to participating hospitals, and published bi-weekly newsletters with hospital performance feedback, success stories, and resources.

The levels of initiatives and partnership activities were fairly uniform throughout the PfP campaign. Without any notable variation in the level of HEN activities, no relationships were observed between intensity or type of initiative or partnership and the changes in performance reflected in the SPC or ITS analyses.

Pressure Ulcers

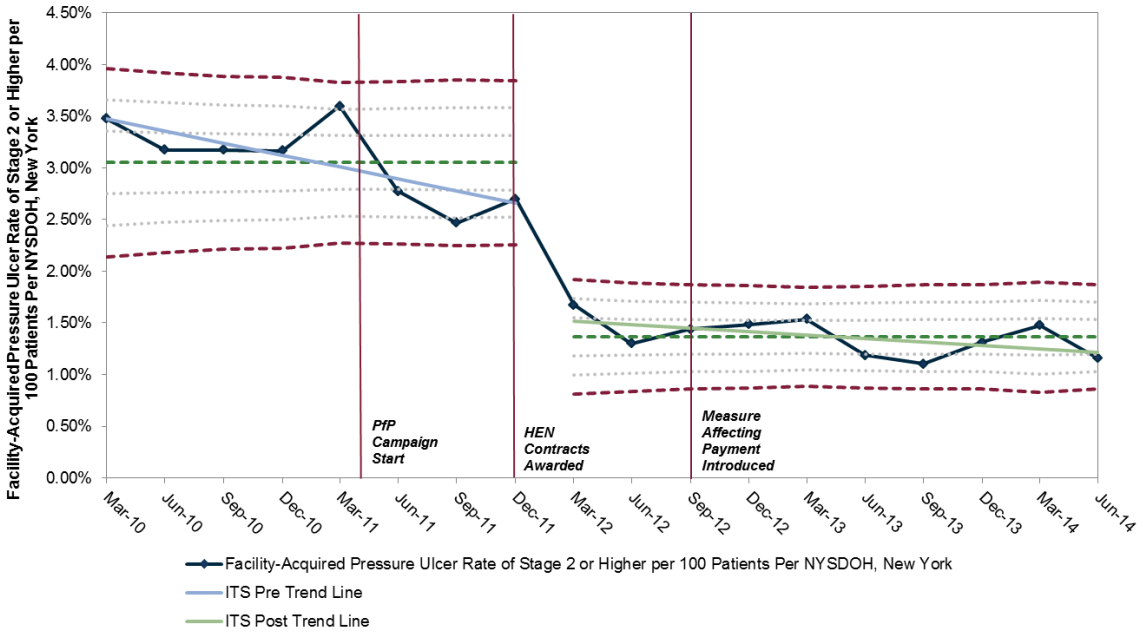
Eleven HENs and HEN cohorts made significant improvements in reducing pressure ulcer rates. The national focus on this type of injury preceded the involvement of the HENs. The first input related to the PfP campaign was a National Priorities Partnership (NPP) webinar on pressure ulcer practices that took place in September 2011, several months after the PfP campaign began, but before HEN contracts were awarded.[5-34] In the summer of 2012, during the HEN component of PfP, a pacing event on the subject was conducted. Another factor that may have influenced results in this area was CMS' institution of payment penalties for hospital-acquired pressure ulcers starting in October 2012.

The New York HEN selected facility-acquired pressure ulcers of stage 2 or higher as one of their target measures in this area. The HEN achieved a 62.02 percent reduction in these pressure ulcers from the baseline rate established in 2011 of 2.89 percent of assessed patients to a 1.16 percent rate in Q2 2014, at the end of the PfP campaign.

As illustrated in Figure 5-21, the SPC chart for New York evidenced a general downward trend from March 2010 through June 2014. It shows evidence of improvement with a shift in the center line just after the HEN contracts were awarded in the first quarter of 2012. The ITS analysis found evidence of a significant structural break toward slowing improvement in January through March 2012. The institution of payment penalties did not appear to affect the gradual downward trend.

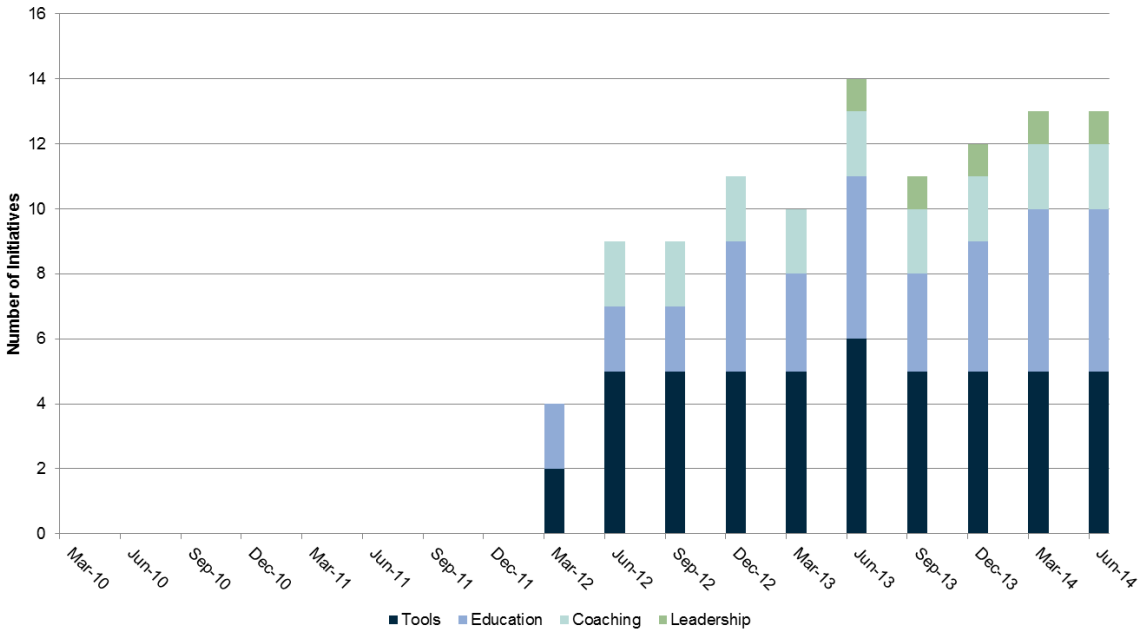
[5-34] The National Priorities Partnership is a partnership of major national organizations gathered under the umbrella of the National Quality Forum as part of the national quality strategy for reducing patient harms. For further information, see http://www.qualityforum.org/Setting_Priorities/NPP/National _Priorities_Partnership.aspx. Accessed on September 23, 2015.

Figure 5-21—Facility-Acquired Pressure Ulcer Rate of Stage 2 or Higher per 100 Assessed Patients, New York



Source: HEN Monthly Reports, November 2014 (n = 58 to 116 hospitals, depending on time period).
Note: Center line and control limits (U' chart) for first phase were calculated with data between January 2010 and December 2011. Center line and control limits (U' chart) for second phase were calculated with data between January 2012 and June 2014.The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits.

Figure 5-22—Pressure Ulcer Initiatives, New York



Source: HEN timeline data fall 2014.
Note: No data available prior to Q1 2012.

The New York HEN established a strong infrastructure which may be responsible for its success in reducing harms in pressure ulcers and in other AEAs, including OB-EED, CLABSI, and CAUTI. The HEN's approach to reducing pressure ulcers followed a systematic set of guiding principles to sustain best practices. They established an infrastructure by creating a website for participating hospitals that contained an overview of the initiative, meeting materials, tools and resources and a data portal for participants to see measure definitions and receive comparative data for analyses. In addition, individual hospitals could post their success stories on the site.

New York did not identify any partnerships that were dedicated specifically to reducing the rate of pressure ulcers. However, the partnership with its QIO, IPRO, to administer the Hospital Survey on Patient Safety Culture annually in 2012, 2013, and 2014 was identified as helpful in the HEN's efforts to reduce patient harms across all AEAs.

The HEN identified several initiatives, both directly related to reducing the incidence of pressure ulcers, and indirectly improving patient care by establishing a culture of safety in participating hospitals. The frequency and types of initiatives are shown in Figure 5-22, above. Beginning in the first quarter of 2012 and continuing monthly through June 2014, the HEN provided a variety of tools and educational initiatives, highlighting best practices to improve the assessment, management and prevention of pressure ulcers. They also shared resources with the New York State Gold STAMP (Success Through Assessment, Management, and Prevention) Program to Prevent Pressure Ulcers.[5-35] Leadership initiatives included a conference in May 2013 where leadership committed to implement safety strategies in the delivery of care through the use of white boards, bedside reporting and multi-disciplinary rounds. The HEN indicated that these techniques were highly effective in reducing pressure ulcers and could be effective in reducing other patient harms.

Information submitted by New York indicated an increase in the frequency of all its initiatives over the course of PfP, adding direct leadership involvement to the mix beginning in Q2 2013 and coinciding with the leveling off of pressure ulcer rates.

VAP

Twelve HENs achieved significant reductions in VAP by the end of PfP. The Ohio Children's Cohort 1 measure VAP Cases per 1,000 Ventilator Days started with one of the higher baselines, a rate of 0.71 cases per 1,000 ventilator days during the year 2011, yet the HEN ultimately surpassed the goal of a 40 percent reduction, reporting a VAP rate of 0.24 cases per 1,000 days in August 2014, a reduction of 47.27 percent.[5-36]

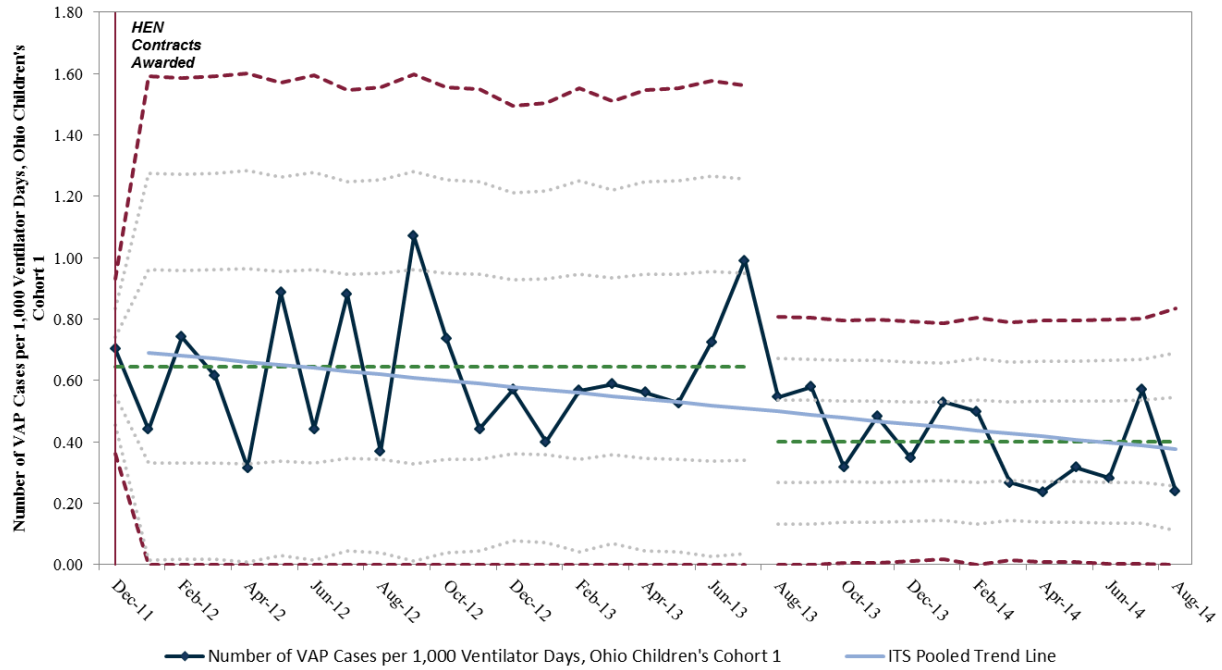
The NCD hosted major pacing sessions devoted to VAP in October 2012 and June 2013. Information provided at the second event demonstrated that of the 12 HENs then participating in VAP, Ohio Children's was one of only two actually exhibiting worsening in their VAP rates. For the HEN to come from that position in June 2013, to surpassing goal by the end of the campaign, was an inspiring achievement. The SPC charts showed a shift in the center line shortly after the HEN's results were made public, achieving a substantial improvement in the next two months. This may be an example of the importance of feedback on performance that was critical to the PfP campaign strategy.

5-35 According to its web site, "The Gold STAMP Program to Reduce Pressure Ulcers is a coalition of organizations convened in New York State to provide resources and education across the continuum of care to improve the assessment, management and prevention of pressure ulcers." (http://www.albany.edu/sph/cphce/goldstamp.shtm). Accessed on September 20, 2015.

5-36 Although the CDC/NHSN changed its definitions of VAP during the PfP time period, those changes did not apply to the pediatric populations treated by the Ohio Children's Hospitals HEN.

The SPC chart for this measure showed evidence of improvement in rate during the PfP campaign, as indicated by a downward a shift in the center line in August 2013, as depicted in Figure 5-23. The ITS analysis did not identify a significant structural break, indicating no sustained departure from the observed downward trend.

Figure 5-23—VAP Cases per 1,000 Ventilator Days, Ohio Children's Cohort 1

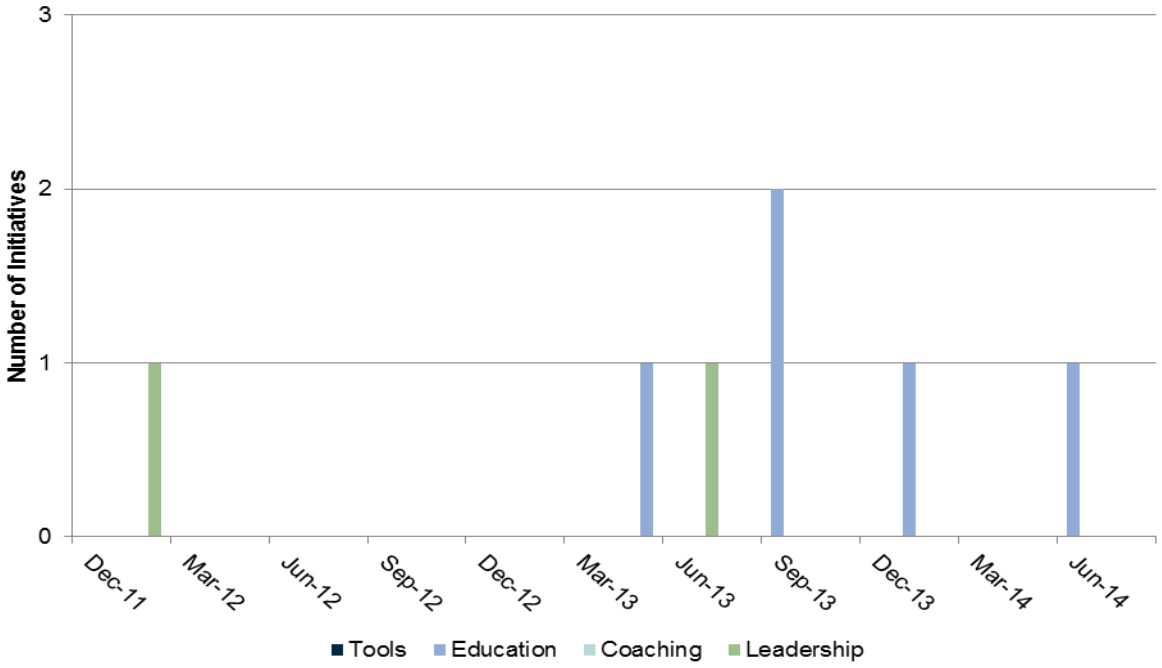


Source: HEN Monthly Reports, November 2014 (n = 25 to 33 hospitals, depending on time period).
Note: Center line and control limits (U' chart) for first phase were calculated with data between January 2011 and February 2013. Center line and control limits (U' chart) for second phase were calculated with data between August 2013 and August 2014.The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. First data point representing CY 2011 was omitted from the ITS analysis, which was restricted to monthly observations.

Although several of the seven initiatives Ohio Children's reported for safety across the board may have impacted VAP, the HEN reported only two initiatives directly addressed the AEA. Most of its broader initiatives were part of Solutions for Patient Safety (SPS), a network of children's hospitals devoted to reducing pediatric inpatient harms.[5-37] Initiatives targeted directly at VAP included leadership gatherings with SPS in February 2012 and July 2013, and educational activities in May and September 2013, and in January and June 2014, covering topics such as patient and family engagement and the importance of providing adequate resources for data collection. There appeared to be little relationship between the timing of these initiatives and the HEN's results for this measure, although the peak in initiatives in October 2013 occurred shortly after a sizable blip in the VAP rate in July 2013, as shown in Figure 5-24.

5-37 This Children's Hospitals' Solutions for Patient Safety is a network of more than 80 children's hospitals working to improve pediatric care across the nation. For more information, see http://www.solutionsforpatientsafety.org. Accessed on September 21, 2015.

Figure 5-24—VAP Initiatives, Ohio Children's



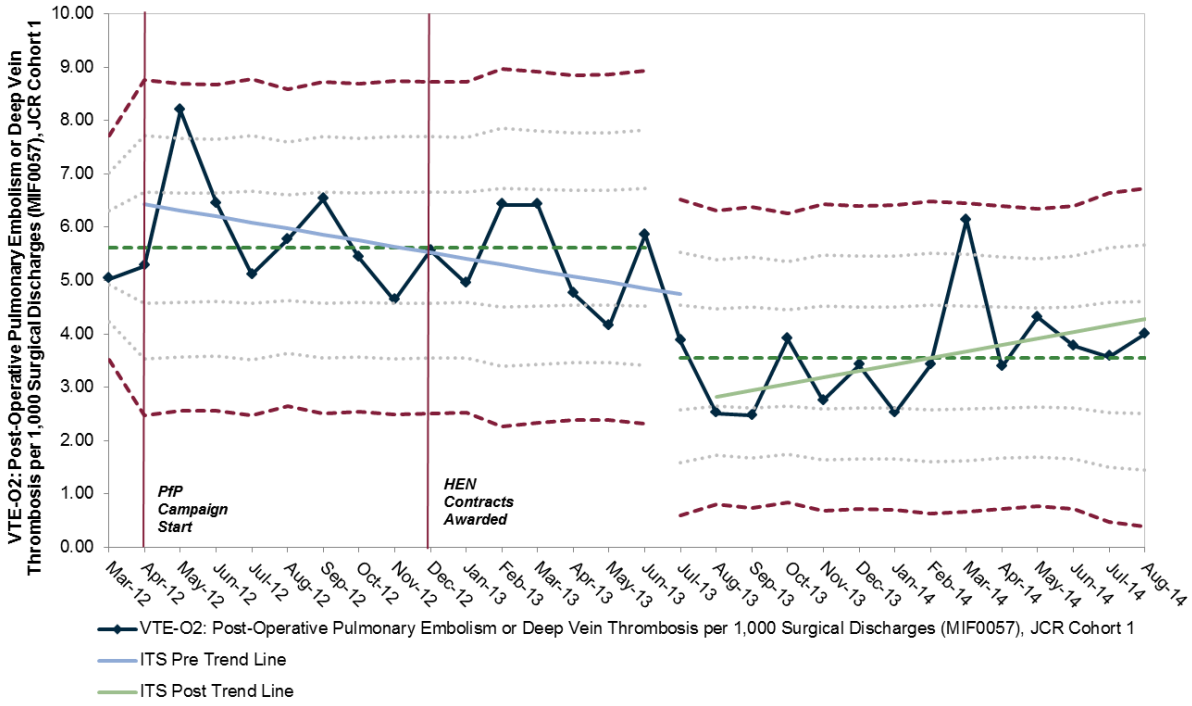
Source: HEN timeline data, fall 2014.
Note: First bar represents baseline period, January to December 2011.

VTE

Among the results from HENs and HEN cohorts that addressed VTE, eleven common measures showed significant improvement in VTE rates while fourteen did not change significantly. JCR Cohort 1's measure Post-Operative Pulmonary Embolism (PE) or Deep Vein Thrombosis (DVT) (VTE) per 1,000 Surgical Discharges, was selected as an example of the work done in this area both to highlight the HEN's excellent work and because it provides as an example of the struggles many of the HENs encountered while trying to maintain their rates of improvement.

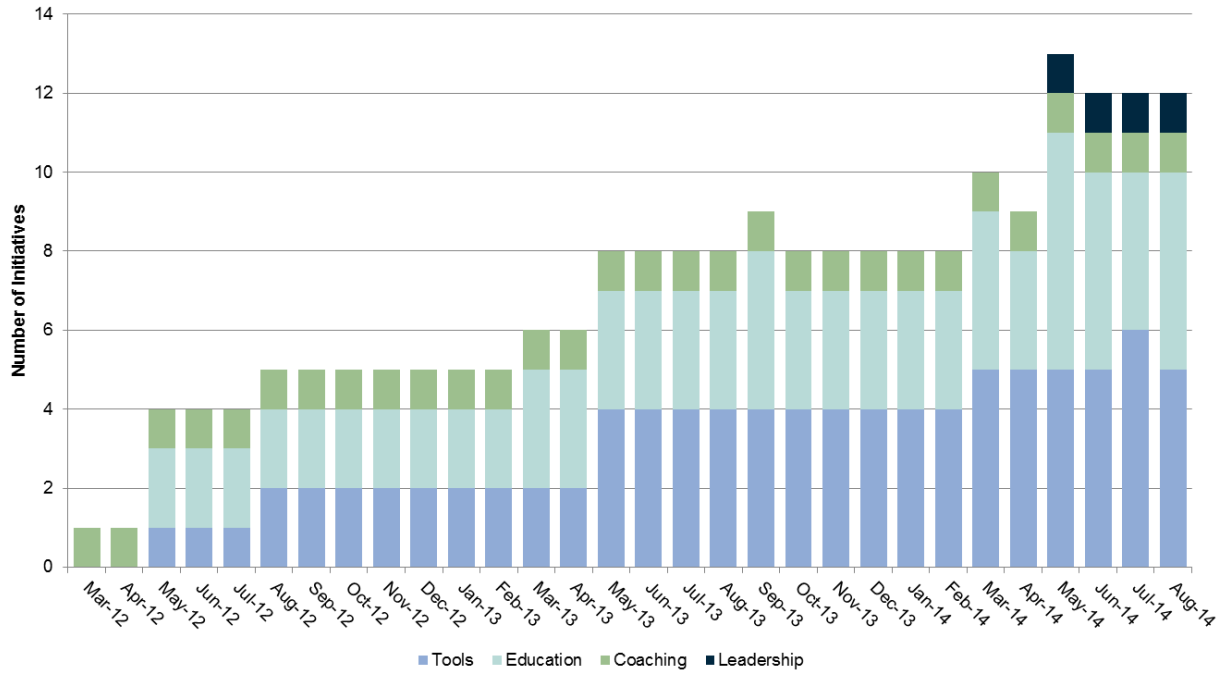
JCR's rate of post-operative VTE for the January 2010 through March 2012 baseline period was 5.03 per 1,000 surgical discharges. Working with 13-24 hospitals, after two months of increase to its highest rate, 8.19 per 1,000 in May of 2012, JCR's rates for VTE trended downward, ending at 4.00 per 1,000 in August 2014. This was a 24.96 percent reduction from baseline, as portrayed in Figure 5-25. The SPC chart showed a downward trend, with evidence of a downward shift in rate in July 2013, when the center line shifted downwards. The ITS analyses showed a structural break the following month, and an upward trend from that point. The HEN appeared to have difficulty maintaining a low level of VTE, since the rate slowly crept up again. Regardless, there was a net improvement by the end of PfP.

Figure 5-25—Post-Operative PE or DVT (VTE) per 1,000 Surgical Discharges, JCR Cohort 1



Source: HEN Monthly Reports, November 2014 (n = 13 to 24 hospitals, depending on time period).
Note: Center line and control limits (U chart) for first phase were calculated with data between January 2010 and June 2013. Center line and control limits (U' chart) for second phase were calculated with data between July 2013 and August 2014. The dashed green line is the center line; the dashed red lines are the upper and lower control limits; the closest dotted lines above and below the center line are the one-sigma limits; and the dotted lines just inside the control limits are the two-sigma limits. First data point representing CY 2010 to Q1 2012 was omitted from the ITS analysis, which was restricted to monthly observations.

Figure 5-26—VTE Initiatives, JCR Cohort 1



Source: HEN timeline data, fall 2014.
Note: First bar represents baseline period, January 2010 to March 2012.

Like many of the HENs, JCR reported taking a comprehensive approach to reducing patient harms cutting across all of the hospital-acquired conditions (HACs). These are summarized in Figure 5-26, above. The HEN characterized their approach as a "kaizen" strategy, eliciting input from employees at all levels of the organization to identify inefficiencies and hazards while focusing continuous daily efforts to improve processes and systems.[5-38] The technique was used in individual hospitals, as kaizen events were scheduled to focus on difficult AEAs, including VTE. "Tracer tools" were created for each AEA to track the care experiences of selected patients while in a participating hospital. The tracer tools provided hospitals with potentially-actionable feedback care provision, treatment, and services based upon the experience of actual patients.

JCR identified two highly-valued partners in its work to reduce VTEs and other hospital acquired conditions. One partnership was with the Northwestern Feinberg School of Medicine, holding collaborative physician calls and developing a physician/leadership change package. The second partnership was with a subcontractor that offered the Agency for Healthcare Quality and Research's (AHRQ's) Hospital Survey on Patient Safety Culture to all of JCR's hospitals.

This approach to continuous improvement, with its initiatives and key partnerships, was associated with excellent results, but only to a certain point. As the rate gradually increased after its low point in September 2013, JCR appeared to have intensified its initiatives, sharing tools, education and coaching in an attempt to turn the trend around. Guest speakers were invited to discuss critical factors for driving down VTE rates, the use of a JCR HEN VTE tracer tool and national trends and prevention methods for VTE. In addition, leadership became more involved in the second and third quarters of 2014. Though the rates worsened

5-38 "Kaizen," a Japanese word indicating "improvement" or "good change," is a label generally applied to the continuous process improvement strategy developed by W. Edwards Deming, Walter A. Shewhart.

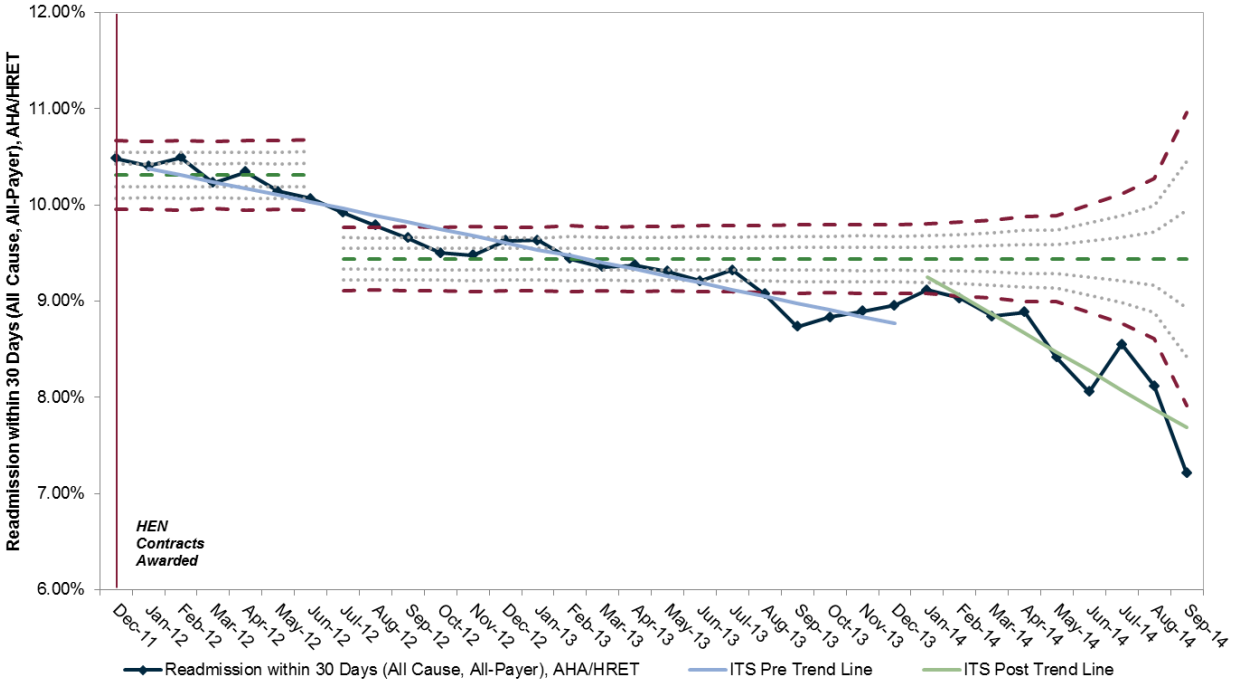
despite the increase in education, coaching, tools and leadership involvement, there may be several different explanations for this result. The impact of various activities may change over time, resources may be moved from one initiative to another, and attention to a particular initiative may diminish. However, given what appear to have been intensive efforts of the HEN to regain its earlier performance, they may have simply reached an effective floor for the patient harm, beyond which further improvement was far more difficult.

Readmissions

The readmissions AEA strongly illustrates the benefits obtained by employing a systematic commitment from leadership and numerous diverse approaches to training (auditory, visual, and kinesthetic) to achieve improvements in national trends. As shown in Chapter 2, the aggregate all-cause, all-payer readmissions rate for the HENs was significantly reduced by 7.34 percent from baseline through July 2014, and the majority of reductions clearly began after the PfP campaign was initiated. In total, twenty seven HENs and HEN cohorts established significant improvements in 30 day readmissions rates from baseline through the end of PfP.

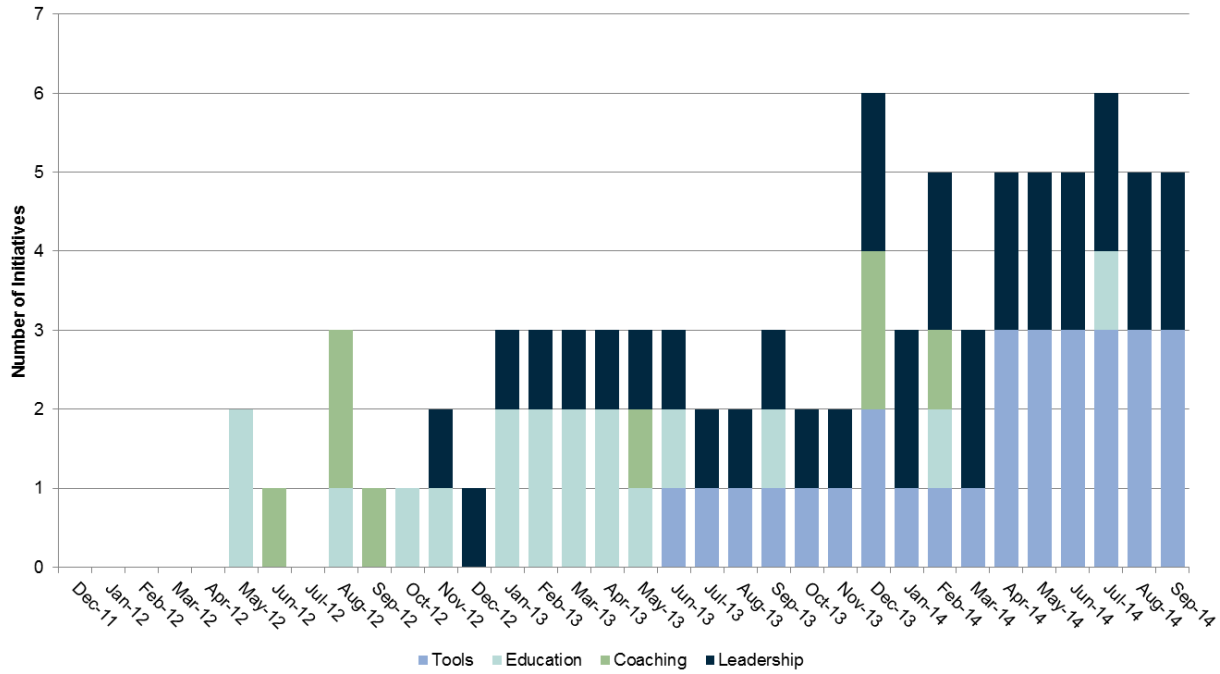
As an example of a high performing HEN, data collected from AHA/HRET Cohort 1 indicates that the participating hospitals succeeded in reducing their readmission rate by 21.24 percent from their baseline rate of 10.48 percent, the average rate for all of 2011. The SPC chart provided in Figure 5-27 shows that readmission rates were trending downward from the time the HEN contracts were awarded in December 2011. There was evidence of improvement, with a shift in the center line in July 2012. There were seven consecutive declines from April 2012 through November 2012. Observations from September 2013 through December 2013 and from February 2014 through September 2014 fell below the lower control limits. The ITS analysis identified a significant break toward a faster rate of improvement in January 2014 continuing through the end of PfP. Similar patterns of improvement were noted in Cohorts 2 and 4, with less improvement in Cohort 3.

Figure 5-27—Readmissions within 30 Days (All Cause, All-Payer), AHA/HRET Cohort 1



Source: HEN Monthly Report, November 2014 (n = 81 to 835 hospitals, depending on the period).
Note: Center line and control limits (U' chart) for first phase calculated using data from December 2011 through June 2012. Center line and control limits (U' chart) for second phase calculated using data from July 2012 through September 2013. First data point representing CY 2011 was omitted from the ITS analysis, which was restricted to monthly observations.

Figure 5-28—Readmissions Initiatives, AHA/HRET Cohort 1



Source: HEN timeline data, fall 2014.
Note: No data available prior to January 2012. For January 2012 and later, no bar is shown when there is no targeted activity or general cross-cutting activity related to this specific measure or its AEA. First bar represents baseline period, January 2011 to December 2011.

Upon examining the type and timing of initiatives reported by AHA/HRET it was apparent that this HEN began its initiatives immediately and continued them throughout the campaign including an effort it described as a "Readmissions ReBoot" in July of 2014. Over 30 different initiatives were documented from May 2012 through August 2014, as displayed in Figure 5-28. The HEN made use of diverse types of training initiatives including regional meetings, boot camps/webinars, monthly coaching calls, change packages and other tools.

One of the early initiatives employed by AHA/HRET to decrease readmissions was a "Data and Coaching Webinar" in early June of 2012. This webinar and others going forward were recorded and archived to be made available to all AHA/HRET HEN participants. All types of hospitals were invited and the webinar began with a background including a discussion on data management strategy, readmission data, preventable readmissions, how to count readmissions, how to compare hospital rates. It included a discussion of best practices in readmissions providing all with a common language and the ability to share stories.

Coaching Calls, Checkpoint calls, webinars, sharing of best practices and readmission tools occurred often, sometimes monthly, throughout PfP. Leadership was involved early on in the PfP campaign, and became increasingly active in 2014, which coincided with the sharp downward trend that occurred during that period. Along with that increase in leadership involvement, starting in the beginning of 2014, AHA documented a strong increase in other activities such as the provision of tools to participating hospitals.

The HEN identified a single and critical partner as the SHA, which provided input into building a multidisciplinary care transition team in the first quarter of 2013. Later in the third quarter of 2014 several mini-collaboratives took place discussing strategies for reducing readmission, providing tools, and presenting case study analyses.

In general, AHA/HRET used diverse approaches to training, used them often, shared ideas and demonstrated a strong commitment from leadership and maintained a vital partnership with its SHA.

Commonalities

These diverse stories highlight a number of elements common to successful approaches to reducing HACs and readmissions. The HENs collaborated with a large variety of organizations of many kinds, the full panoply of programs and partners recruited by PfP leadership, plus a number of partners identified on a state or local level. A rich pattern of cross-insemination of ideas and approaches is evidenced by nearly all of these HENs in nearly all of these AEAs.

These HENs made great use of the learning community available to them through PfP, using its websites and regular events for communicating approaches, brainstorming solutions, and evaluating performance. Several different management techniques were used as a framework for improvement, most notably kaizen, and lean 6 Sigma. Both techniques were used by the HENs that generated the results discussed above.

All of these factors were seen as crucial to PfP's mission, and were actively encouraged throughout the campaign. The HEN timeline data reflect broad participation in the programs provided by PfP leadership, and demonstrate the breadth of initiatives, pacing events, and feedback to continuously improve performance.

While many of the tactics associated with the observed reductions in patient harms had been used before in different hospitals or regions, PfP seems to have provided a dynamic platform for exposing many hospitals in the country to a fairly unified message regarding best practices in these areas of interest which can be seen as changing the standard of care for not only the HAC and readmissions areas, but for a broader approach to quality improvement.

6. Estimation of Costs Averted

One important effect of the reduction in harms is the cost savings associated with not having to treat those harms. The Evaluation Contractor had two approaches to estimating cost averted. The first approach is to measure the costs adverted from the national reductions in harms in the focused area of the Partnership for Patients (PfP) campaign. These reductions in costs are the result of all the different private and public efforts undertaken to reduce harms and cannot be directly attributed to the PfP campaign. The Evaluation Contractor then estimates the impact of Hospital Engagement Network (HEN) alignment on Medicare expenditures.

Estimates of Averted Costs from National Reduction in Adverse Events Methodology

The harm reductions described in Chapter 2 are estimated to have saved at least $8.67 billion. This is the lower of two different sets of estimates of the cost savings associated with the national reductions in adverse events available from two different sources. Each set of estimates was derived using different methods and data sources and included different adverse events and PfP focus areas. One set was developed by the Evaluation Contractor and the other by the Agency for Healthcare Research and Quality (AHRQ). Due to data availability issues, neither of the estimates covers all the PfP focus areas and neither covers the full PfP time period for all the harm areas.[6-1] In this section, the Evaluation Contractor describes the methods used to obtain each of these measures.

The Evaluation Contractor has estimated that harm reductions nationally have resulted in at least $8.67 billion in cost savings between 2010 (the year before PfP began) through either 2013 or mid-to-late 2014 (varying by measure). In this conservative estimate, many harms that were known to occur were not included in the cost savings estimate because there were no cost-per-event data from the literature that were well-matched to the type of harm being counted, so that the exclusions included adverse drug events (ADEs), stage 1 pressure ulcers (more severe pressure ulcers were counted), and harms that did not fall into the PfP focus areas. The estimate from AHRQ (which is less conservative) suggests savings may total $11.98 billion through 2013.[6-2]

AHRQ's Estimate of Reductions in Cost Associated with Fewer Adverse Events

The AHRQ estimate of reductions in cost associated with fewer adverse events is the sum of the estimated cost savings for 2011, 2012, and interim final 2013. AHRQ's method is to subtract the national estimate of the number of adverse events in each follow-up year (2011, 2012, and 2013) from the number estimated for 2010. The difference for each year is summed and multiplied by the estimated cost per event, which was developed prior to the start of the PfP by an interagency United States (U.S.) Department of Health and Human Services (HHS) group, including representatives from the Centers for Medicare & Medicaid Services (CMS), AHRQ, the Centers for Disease Control and Prevention (CDC), and other agencies.

The AHRQ National Scorecard Data used in the estimates provide approximations of the incidence of adverse events for 2010, 2011, 2012, and 2013, mostly drawn from a dataset known as the Medicare Patient Safety Performance Monitoring System (MPSMS), which is based on a nationwide purposeful sample of

6-1 CMS intends to contract for an additional evaluation to incorporate data from 2014 not available in time to complete analysis for this report.

6-2 In December 2015, HHS announced an updated AHRQ interim estimate through 2014 of $20 billion in savings, see the following website location for details on the estimate: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2014.html.

inpatient charts, CMS Inpatient Quality Reporting (IQR) sample. In addition, the Scorecard includes estimates of surgical site infections (SSIs) from CDC surveillance data, and estimates of AHRQ Patient Safety Indicators (PSIs) for obstetrical events (2 PSIs) and several events in the "all other hospital-acquired conditions (HACs)" category (4 PSIs) derived from all-payer claims data (Hospital Cost and Utilization Project [HCUP] data). The data are presumed generalizable to the full U.S. population.[6-3] Table 6-1 provides the count of adverse events averted during 2011-2013, the cost estimate per adverse event used, and estimated costs averted. More detail is provided in Appendix D, including sources for the AHRQ cost per event estimates and the breakdown of events averted by year.

Table 6-1—AHRQ Estimate of Costs Averted Due to National Decrease in Harm

PfP Harm Area	Cost per Adverse Event (See Appendix D Table X for Sources)	2010 Count of Adverse Events (Rounded)	Reduction in Adverse Events During 2011-2013, Using 2010 Baseline (Rounded)	Reduction in Costs During 2011-2013, Using 2010 Baseline
ADE	$5,000	1,621,000	577,000	$2,885,000,000
CAUTI	$1,000	400,000	190,000	$190,000,000
CLABSI	$17,000	18,000	10,800	$183,600,000
Falls	$7,234	260,000	50,000	$361,700,000
OB-Adverse Events	$3,000	82,000	10,000	$30,000,000
Pressure Ulcers	$17,000	1,320,000	280,000	$4,760,000,000
SSI	$21,000	96,000	45,000	$945,000,000
VAP	$21,000	38,000	8,000	$168,000,000
VTE	$8,000	28,000	5,000	$40,000,000
All Other HACs	$17,000	894,000	142,000	$2,414,000,000
Totals (Based on Unrounded Numbers)		**4,757,000**	**1,317,800**	**$11,977,300,000**

Source: AHRQ publications available at http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html and http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2013.html.

6-3 Although the data are from a sample, statistical testing has not been performed by AHRQ and results are not available to the Evaluation Contractor at this time.

Evaluation Contractor's Estimation of Cost Savings

The Evaluation Contractor developed a second estimate of cost savings, which has both similarities and differences from the AHRQ estimate, as shown in Table 6-2.[6-4] Where the approaches differ, the table also displays the Evaluation Contractor's rationale for choosing an alternative method.

Table 6-2—Major Differences Between AHRQ and the Evaluation Contractor Estimates

Aspect of the Estimate	Different Between the Evaluation Contractor and AHRQ	Evaluation Contractor's Rationale for Difference
Cost per event averted for each area	Yes–For some areas of harm	The Evaluation Contractor uses an original analysis of the HCUP cost data, as well as a recent literature review not available when AHRQ created its estimation method
Inclusion/non-inclusion of obstetrical early elective deliveries (OB-EED)	Yes—Evaluation Contractor includes OB-EED, AHRQ does not include this	The Evaluation Contractor has data on OB-EEDs which are not included in the AHRQ estimate
Inclusion of ADEs and "all other HACs"	Yes—AHRQ includes this, but the Evaluation Contractor does not	The Evaluation Contractor finds the cost-per-event estimates for these two areas to be for events that are different enough from those measured in the data on the number of events that creating an estimate of costs averted by multiplying the cos-per-event estimates with the very differently-measured data on number of events is not credible
Method of calculating events averted	Yes—AHRQ estimates the national number of events each year, normalized to a constant number of discharges, and subtracts the baseline number, whereas the Evaluation Contractor calculates Actual–Expected as described below	The Evaluation Contractor's method is needed when varying denominators are used that change over time; AHRQ's method works for its purposes because the denominator is normalized to the number of adult discharges nationally
Measures used	Yes—different for 3 areas: falls, venous thromboembolism (VTE), OB harm other than OB-EED	The Evaluation Contractor uses narrower measures for falls (falls with fracture) and pressure ulcers (stages 2+) because these more closely match the data used to create the cost-per-event estimate—no cost-per-event estimates exist approximating the broader measures used by AHRQ to count events. For OB harm, the Evaluation Contractor includes injury to neonate. The Evaluation Contractor uses its own estimates for OB harm other than OB-EED and VTE because of improved ability to match with updated cost-per-event information.

Source: The Evaluation Contractor's comparison of the components and results of the two estimates.

6-4 The cost averted estimates provided here by the Evaluation Contractor were evaluated by an actuary with Mercer and recommendations to improve the cost estimates were incorporated into the current methodology.

Methods for Calculation of Adverse Events

- The basic calculation of numbers of adverse events averted (for a given time period of interest) for the areas of falls, OB-EED, OB-Other, readmissions, and VTE is as follows:
  - *Number of Adverse Events Averted* is the difference between the *Number of Adverse Events Expected* (i.e., projected) and *Actual Number of Adverse Events Observed*, where the number of adverse events expected is calculated by multiplying the adverse event rate from a "baseline" or pre-PfP period by a denominator from the current period of interest. The denominator measures patient volume (e.g., number of patient discharges or inpatient days). Intuitively, this calculation compares the number of adverse events actually observed during the period of interest to the number of events that would have occurred had the event rate remained unchanged from the previous time period.
- The method used by AHRQ to calculate numbers of adverse events averted was described above and is incorporated in the Evaluation Contractor's use of AHRQ's estimates for the areas of SSI, catheter-associated urinary tract infection (CAUTI), central line-associated bloodstream infection (CLABSI), ventilator-associated pneumonia (VAP), falls, and pressure ulcers, except that the AHRQ estimates of falls and pressure ulcers were adjusted to better match available cost-per-event data. Specifically, for pressure ulcers the Evaluation Contractor subtracted an estimated number of stage 1 pressure ulcers (estimated by the Evaluation Contractor based on National Database of Nursing Quality Indicators® (NDNQI®) data, so that the severity of pressure ulcers would be more consistent with the available cost estimate, which excluded stage 1.[6-5] For falls, the Evaluation Contractor multiplied AHRQ's estimate for all falls averted by the estimated percent of falls that are fractures, using AHRQ's own data for 2011.
- Where data were limited to a specific payer population (Medicare), the Evaluation Contractor's cost savings analysis used historical ratios of adverse events among Medicare patients to events among non-Medicare patients to create an all-payer averted events number. This extrapolation was used for VTE.

Data Sources for Estimates of Averted Events

The Evaluation Contractor used the following sources for estimates of averted events in this report:

- For CAUTI, CLABSI, VAP, falls, and pressure ulcers (though falls and pressure ulcers were adjusted as described below), the Evaluation Contractor cost savings estimate used the AHRQ estimates of events averted between the baseline of 2010 and 2013, which were based on the MPSMS medical chart reviews and extrapolation methodology. For more details, see http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html.
- For SSI, the Evaluation Contractor cost savings estimate used AHRQ estimates of events averted between the baseline of 2010 and 2013, which were in turn drawn from national estimates made by the CDC for a set of 17 common surgical procedures.
- The sources of data used to calculate the number of adverse events averted for the other areas are as follows (the data sources are further described in Appendix B):
  - HEN-submitted data were used to calculate averted obstetrical adverse events other than OB-EED (OB-Other).
  - Medicare claims data processed by CMS' Health Policy and Data Analysis Group were used to calculate averted VTE events.

[6-5] NDNQI® is a registered trademark of the American Nurses Association (ANA). NDNQI® data were supplied by ANA. The ANA disclaims responsibility for any analyses, interpretations, or conclusions.

  - National Vital Statistics System (NVSS) data were used to calculate number of OB-EEDs averted.
  - For readmissions averted, the data were from AHRQ for Medicare and non-Medicare discharges, derived from the Nationwide Readmissions Database, which is based on HCUP State Inpatient Databases and MedPAR.
- Given the variation in date ranges in the HEN-submitted data, there are multiple options for selecting the periods to be deemed "baseline." Baseline date ranges were chosen to ensure consistency with the following principles:
  - The baseline should be as close to 2010 as possible, since this is the official PfP baseline.
  - The time period of the baseline should be as close to one full year as possible, since quarterly data are likely to be less stable.
  - The baseline date range should avoid time periods where data issues are known to have been severe. In particular, Medicare claims for periods prior to Q2 2011 were characterized by poor present-on-admission (POA) reporting, on which accurate calculation of AHRQ PSIs depends, and/or use only up to 9 diagnosis codes (vs. up to 25 beginning in 2011).
  - If there are no potential baseline periods that meet the above criteria, then available HEN-submitted data are used.

Sources for Cost Per Event

Appendix D shows the estimates of cost per event that were used, based on a literature review by the Evaluation Contractor in 2012–2013, updated with a more recent synthesis article (Zimlichman et al. 2013) and an analysis by the Evaluation Contractor of cost data from the HCUP in 2014.

Given that Zimlichman et al. represented a systematic literature review and meta-analysis, the Evaluation Contractor used its estimates of per-event cost for CAUTI, CLABSI, SSI, and VAP. Also, where possible, the Evaluation Contractor prioritized use of the estimates created by the Evaluation Contractor's original analysis of HCUP data for VTE and OB harms other than OB-EED. Estimates of OB-EED cost were not available from the literature, so the Evaluation Contractor created an estimate based on key facts contained in the literature (Table 6-3). The most recent measure found in literature sources that was broadly measured and matched as closely as possible to the data for events averted was used for readmissions and falls.

Estimates were updated to account for inflation using the Consumer Price Index - Urban Consumers (CPI-U) from the year of the study data that were available. All estimates are in 2014 dollars.

The HCUP analysis provided original cost-per-event estimates for Medicare and non-Medicare populations for VTE (PSI-12), and all-populations measures for the three OB-Other harm measures (PSIs-17, -18, and -19). The estimates are the difference in hospital costs (billed charges, adjusted by hospital cost-to-charge ratios) during an index stay between patients with and without an adverse event, using a matched sample from the 2009 to 2011 HCUP data that were available at the time of analysis for 12 states.

Table 6-3 provides the Evaluation Contractor's estimate of the number of events averted and the time frame and source for the estimate, and Table 6-4 provides the AHRQ and Evaluation Contractor's estimates by area and in total, side by side, with comments on possible explanation of the differences.

Table 6-3—Number of Estimated Events Averted, Time Frames, and Data Sources in the Evaluation Contractor's Estimate

Adverse Event Area	Cost Per Adverse Event (See Appendix D for Sources)	Estimated Number of Events Averted	Time Frame and Data Source for Number of Events Averted	Estimate of Total Costs Averted
CAUTI	$989	190,000	2010 baseline, 2011–2013 estimate, AHRQ's MPSMS	$187,910,000
CLABSI	$50,568	10,800	2010 baseline, 2011–2013 estimate, AHRQ's MPSMS	$546,134,000
OB-EED	$9,762 for OB-EEDs resulting in neonatal intensive care unit (NICU); the number of NICU stays estimated to equal 0.0996 x the number of OB-EEDs[a]	118,103	2010 baseline, events averted 2011–2013. Data from the NVSS (vital records data) for most U.S. counties.	$114,831,000[a]
Falls with Fracture	$12,965	2,070	2010 baseline, 2011-2013 estimate, AHRQ's MPSMS, adjusted to represent falls with fracture (estimated to represent 4.1 percent of all falls[c])	$26,838,000
Pressure Ulcers (excludes stage 1)	$12,565	218,400	2010 baseline, 2011-2013 estimate, AHRQ's MPSMS, adjusted to exclude stage 1 pressure ulcers (estimated to represent 22 percent of the AHRQ count of averted pressure ulcers[b])	$2,744,196,000
PSI-12—VTE	$17,666 (Medicare) $27,691 (Non-Medicare)	5,718	Q2-Q4 2011 baseline, events averted 2012–Q4 2014. Medicare claims data, extrapolated to national estimate using assumption from the literature.	$126,600,000
PSI-17—Injuries to Neonate	$1,145	259	HEN data submitted November 2014, including approximately 1,300 hospitals. Mixed baselines and follow-up periods. Events averted estimated HEN by HEN and summed, with rules for ensuring rough comparability of the population over time. Most follow-up periods ending Q1 2014 or later.	$297,000

Table 6-3—Number of Estimated Events Averted, Time Frames, and Data Sources in the Evaluation Contractor's Estimate				
Adverse Event Area	**Cost Per Adverse Event (See Appendix D for Sources)**	**Estimated Number of Events Averted**	**Time Frame and Data Source for Number of Events Averted**	**Estimate of Total Costs Averted**
PSI-18—Obstetric Trauma-Vaginal Delivery with Instrument	$114	2,670	HEN data submitted November 2014, including approximately 1,500 hospitals. Mixed baselines and follow-up periods. Events averted estimated HEN by HEN and summed, with rules for ensuring rough comparability of the population over time. Most follow-up periods ending Q1 2014 or later.	$304,000
PSI-19—Obstetric Trauma-Vaginal Delivery without Instrument	$197	9,844	HEN data submitted November 2014, including approximately 1,600 hospitals. Mixed baselines and follow-up periods. Events averted estimated HEN by HEN and summed, with rules for ensuring rough comparability of the population over time. Most follow-up periods ending Q1 2014 or later.	$1,939,000
SSI	$22,942	45,000	2010 baseline; 2011–2013 estimate, CDC's estimate provided by AHRQ.	$1,055,332,000
VAP	$44,310	8,000	2010 baseline, 2011–2013 estimate, AHRQ's MPSMS.	$355,280,000
Readmissions	$15,477 (Medicare) $13,311 (Non-Medicare)	222,125	2010 baseline, 2011-2013 estimate, data based on 1) Nationwide Readmission Database (NRD) developed by AHRQ from the HCUP State Inpatient Databases; and 2) Medicare Provider Analysis and Review (MedPAR) files from CMS. Data provided by AHRQ staff member Joanna Jiang, August 27, 2015.	$3,502,998,000

Source: Evaluation Contractor's analysis of data sources listed.

[a]The cost per event, and thus the total estimate of costs averted from OB-EED declines may overestimate savings from OB-EED reduction; they are based on the relationship between OB-EEDs and NICU stays found in the literature, however, this relationship did not appear to hold at the county level for analyses conducted with NVSS data (discussed in Chapter 4). Even if true savings from the decrease in OB-EEDs was zero, the total costs averted estimate remains conservative given the large number of events counted by AHRQ for which there is no closely corresponding cost-per-event estimate, therefore they were omitted entirely.

[b]The stage of pressure ulcers was not available for a majority of pressure ulcers in the MPSMS data, so the percentage likely to be stage 1 was estimated by the Evaluation Contractor based on the NDNQI pressure ulcers data for Q1 2013, for which the Evaluation Contractor had complete data on stages.

[c]The percentage of falls that are falls with fracture was based on the MPSMS data for 2011.

Table 6-4—Cost Reductions Estimated by AHRQ and by the Evaluation Contractor			
PfP Focus Area	**AHRQ Estimate of Cost Reductions Due to Reductions in Adverse Events from the AHRQ National Scorecard During 2011-2013 Compared with 2010 (Rounded)**	**Estimate of Cost Reductions Due to Reductions in Adverse Events from the Evaluation Contractor[a] (Rounded)**	**Comments on Sources of Difference for the Two Estimates**
Cost Reductions Estimated by Both AHRQ and the Evaluation Contractor: (Estimates are National in Scope Unless Noted as Sub-National Due to Data Source Limitations)			
VTE	$40,000,000	$126,600,000	Per-event cost used by the Evaluation Contractor per analysis of HCUP data is more than double that of AHRQ's cost per event. Estimated number of events averted is roughly similar (5,000 for AHRQ vs. 4,115 for the Evaluation Contractor).
VAP	$168,000,000	$355,280,000	Per-event cost used by the Evaluation Contractor per literature is nearly double that of AHRQ's cost per event. Both use AHRQ's number of events averted.
CAUTI	$190,000,000	$187,910,000	Per-event cost is roughly similar. Both use AHRQ's number of events averted.
CLABSI	$183,600,000	$546,134,000	Per-event cost used by the Evaluation Contractor per literature is nearly triple that of AHRQ's cost per event. Both use AHRQ's number of events averted.
Falls	$361,700,000	$26,838,000 (sub-national)	Per-event cost estimate used by AHRQ is more than 13 times the estimate used by the Evaluation Contractor. AHRQ's per-event cost estimates costs for falls with injury although AHRQ counts all falls in its events count. Number of events averted is much smaller for the Evaluation Contractor because only reported falls with fracture are captured, in order to match the available cost per event number.

Table 6-4—Cost Reductions Estimated by AHRQ and by the Evaluation Contractor			
PfP Focus Area	**AHRQ Estimate of Cost Reductions Due to Reductions in Adverse Events from the AHRQ National Scorecard During 2011-2013 Compared with 2010 (Rounded)**	**Estimate of Cost Reductions Due to Reductions in Adverse Events from the Evaluation Contractor[a] (Rounded)**	**Comments on Sources of Difference for the Two Estimates**
OB-Other	$30,000,000	$2,540,000 (sub-national)	Per-event cost estimate used by AHRQ is much, much higher than the HCUP analysis-based estimates used by the Evaluation Contractor ($3,000 per event vs. $108, $185, and $1,080 depending on type of event[b]). Number of events averted in the Evaluation Contractor's number is about 28 percent higher than that of AHRQ, due to inclusion of injury to neonates in the measure and use of a different data source, which extends beyond 2013, although it includes only hospitals that reported the measures to their HENs.
Pressure Ulcers	$4,760,000,000	$2,744,196,000	Per-event estimate by the Evaluation Contractor is lower than that used by AHRQ, because the AHRQ estimate is an estimate of cost for severe pressure ulcers, whereas the Evaluation Contractor's cost estimate includes stage 2 pressure ulcers, which are less severe. Number of events averted is higher in the AHRQ estimate because AHRQ counts all pressure ulcers; the Evaluation Contractor excludes stage 1 pressure ulcers in order to match the available cost-per event estimate, which is based on pressure ulcers stages 2 and higher.
SSI	$945,000,000	$1,055,332,000	Per-event cost estimate is similar in both estimates (updated for inflation in the Evaluation Contractor's estimate). Both estimates use the AHRQ estimate of events averted (based on the CDC's estimates).

Table 6-4—Cost Reductions Estimated by AHRQ and by the Evaluation Contractor			
PfP Focus Area	**AHRQ Estimate of Cost Reductions Due to Reductions in Adverse Events from the AHRQ National Scorecard During 2011-2013 Compared with 2010 (Rounded)**	**Estimate of Cost Reductions Due to Reductions in Adverse Events from the Evaluation Contractor[a] (Rounded)**	**Comments on Sources of Difference for the Two Estimates**
Cost Reductions Estimated only by the Evaluation Contractor			
OB-EED	Not estimated	$114,831,000	
Readmissions	Not estimated	$3,502,998,000	
Cost Reductions Estimated only by AHRQ			
ADEs	$2,885,000,000	Not estimated	
All Other HACs	$2,414,000,000	Not estimated	
Total	**$11,977,300,000**	**$8,674,830,530**	

Source: Column 1 estimates are from Noel Eldridge, AHRQ Center for Quality Improvement and Safety, updated on August 26, 2015. Column 2 estimates are from the Evaluation Contractor, using methodology described above.
Note: Estimates for each harm area were rounded. The total differs from the sum of the estimates because the total was derived using the sum of unrounded data.
[a]Estimates in this column for OB-EED, falls, OB-Other, pressure ulcers, and VTE are partial estimates limited to available data as follows: OB-EED estimates reduced costs due to estimated reduced use of the NICU only; falls, and OB-Other estimates cover only hospitals reporting to their HEN; pressure ulcers covers only stage 2 and higher pressure ulcers; and VTE baseline data are for Q2–Q4 2011 rather than 2010, due to data issues in earlier data.
[b]The AHRQ estimate of $3,000 included costly types of harms not likely included in the PSI data that are the basis for the count of events.

Estimation of Costs Averted As a Result of the HEN Component of the PfP Campaign

As noted earlier, a key outcome of the PfP campaign is to reduce those costs associated with adverse events. The decrease in costs may result from the targeted harm reductions, but it can also result from avoiding other harms if aligning with a HEN led to hospital-wide changes in culture or practices. It is possible that the impact analyses summarized in earlier chapters may not be able to detect such discrete changes in harms or readmissions if any individual change is small. As such, focusing on changes in overall healthcare expenditures, which would include reductions due to prevention of non-targeted as well as targeted adverse events, may allow for the detection of more nuanced effects of HEN alignment.[6-6]

In addition to healthcare expenditures incurred during an initial (index) hospital admission being examined, it is also important to measure expenditures for all healthcare services received in the post-discharge period. Although adverse events can involve substantial expenditures during the initial hospital stay, they may also create non-trivial healthcare costs following discharge, which are important to capture. For example, a fall may lead to substantial rehabilitation costs, which would be missed if only expenditures during the hospital stay were measured. The consequences of adverse events may affect the need for different health services. The Evaluation Contractor examined individual types of Medicare expenditures in the post-discharge period, including outpatient, inpatient (e.g., readmissions), and Durable Medical Equipment (DME).

6-6 HEN alignment would be expected to unambiguously lead to a reduction in costs if it reduced adverse events but had no effect on mortality rates or index admissions. If HEN alignment reduced the hospital mortality rate, it is possible that medical expenditures per index discharge could increase. Similarly, if HEN alignment reduced index admissions among those less at risk for an event, then it is possible that medical expenditures per index discharge could increase due to change in the composition of patients. As such, the total effect of HEN alignment on medical expenditures depends on the relative extent to which adverse harms, deaths, and index admissions were affected.

To account for the possibility that the campaign improved care for all patients and not just for those at risk for particular adverse events, the analyses focus on all individuals with a hospital admission during the study period, not just those at risk for a particular adverse event.

Medicare Expenditures Data

For this analysis, the Evaluation Contractor used Medicare claims data to examine expenditures of all fee-for-service (FFS) Medicare beneficiaries with an acute care hospital admission between January 2009 and December 2013.[6-7,6-8] As discussed, expenditures during the initial index hospital admission as well as in the post-discharge period are analyzed. Table 6-5 provides a breakdown and an explanation of the different types of Medicare expenditures examined in this chapter. In the analyses, all expenditures are adjusted for price-inflation.[6-9] Further details on the identification of index discharges and the construction of post-discharge expenditures can be found in Appendix D.

Table 6-5—Description of Expenditures Measures

Measure	Description
Total Expenditures[6-10]	Combined Medicare expenditures from index admission and from post-discharge period.
Index Admission	Medicare expenditures associated with index admission, including both hospital and physician payments.
Post-Discharge[6-11]	Total Medicare expenditures in the post discharge period, including inpatient and outpatient expenditures and expenditures for DME, home health services, and skilled nursing facilities (SNF), hospice, and medical professionals.
Post-Discharge Inpatient (IP)	All inpatient Medicare expenditures in the post discharge period, regardless of the inpatient setting (e.g., acute care hospitals, long-term care hospitals, and rehabilitation hospitals).
Outpatient	All outpatient Medicare expenditures from the post-discharge period, such as emergency department (ED) expenditures, observation stay expenditures, and other outpatient amounts.
DME	Medicare expenditures from the post-discharge period for claims by DME suppliers.
Home Health (HH)	Medicare expenditures from the post-discharge period on home health services, such as intermittent skilled nursing care, physical therapy, speech-language pathology services, and continued occupational services.
SNFs	Medicare expenditures from the post-discharge period for care provided at SNFs.
Hospice	Medicare expenditures from the post-discharge period for care by hospice providers.

6-7 Only beneficiaries covered by Medicare FFS Parts A during the entire index stay period (index stay admission date through index stay discharge date) and FFS Parts A and B throughout the duration of the lookout period (either 90- or 180-days from the time of the index stay discharge) are included.

6-8 Types of hospitals considered for identifying the index inpatient discharges include only short-term acute hospitals. See Appendix D for more details.

6-9 Expenditures are adjusted for price inflation using the monthly Consumer Price Index for All Urban Consumers (CPI-U) by census region, where January 2010 is the base time period.

6-10 Total expenditures are derived by summing the expenditures from the index admission and from the post-discharge period.

6-11 Post-discharge expenditures are derived by summing Medicare expenditures on post-discharge inpatient admissions, outpatient care, DME, HH services, SNF care, hospice services, and medical professionals in the post-discharge period.

Table 6-5—Description of Expenditures Measures	
Measure	**Description**
Professional Services (physician, physician assistants, nurse practitioners, and other providers paid under Part B.)	Medicare expenditures from the post-discharge period for care provided by healthcare professionals such as physicians, physician assistants, and nurse practitioners.

Note: These expenditures breakdowns are derived from Medicare FFS claims. In the analyses, all expenditures are adjusted for price inflation unless otherwise noted.

Defining the Post-Discharge Period

Two post-discharge periods are used in the analyses in order to examine whether the difference-in-differences results are sensitive to the time period analyzed: 90 days and 180 days after the index discharge. The 90-day window after a hospital discharge has been used in other studies focusing on medical care expenditures (Encinosa and Hellinger 2008; Kandilov, Coomer, and Dalton 2014). A window of 180 days is also examined to assess the sensitivity of the results in allowing a more generous post-discharge period. If these lookout periods are not long enough to capture the full effect of PfP on costs, then the difference-in-differences estimates will be lower bounds of the true estimates. Conversely, if these windows are too long, then the difference-in-differences estimates will be unbiased since the comparison group is measured over the same period, with the same window, as the HEN aligned group.

Estimating the Impact on Medicare Expenditures

In this chapter, difference-in-differences estimates are derived from comparisons of Medicare expenditures for patients with an index hospital admission at a HEN-aligned hospital to those of patients at non-HEN-aligned hospitals both before and after the start of PfP. As such, this is a discharge-level empirical analysis. The difference-in-differences estimates represent the impact of HEN alignment on costs for calendar year 2013.[6-12] In constructing the comparison group of non-HEN-aligned hospitals, propensity score weights were applied to give higher weights to those non-HEN-aligned hospitals with observable characteristics more similar to HEN-aligned hospitals. Details on the construction of the propensity score weights and the difference-in-differences analyses can be found in Appendix D.

[6-12] Standard errors (SEs) are always clustered at the hospital level. Further details on the difference-in-differences model and the set of patient controls used in the analysis are provided in Appendix D.

Did HEN-Alignment Affect Medicare Expenditures?

This analysis did not identify a difference in Medicare expenditures from index admissions incurred in HEN-aligned versus non-HEN-aligned hospitals. Table 6-6 and Table 6-7 show the difference-in-differences impact estimates from the analysis. HEN alignment has no statistically significant effects for both the 90- and 180-day outlook period. Examining the different categories of expenditures, such as post-acute care, professional services, and DME costs, the Evaluation Contractor finds similar results showing that expenditures per discharge were not affected by HEN alignment. The estimates are fairly precise: the size of the coefficients and their standard errors (SEs) are small, particularly relative to the mean expenditures shown in the first columns.[6-13] This suggests that it is not a lack of statistical power that is driving the insignificant results. Robustness checks are provided in Appendix E, where the analysis is repeated using 2010 as the baseline year (rather than 2011); nominal expenditures (i.e., not price-adjusted); collapsed time periods (i.e., pre-PfP vs. post-PfP), and finer expenditure categories. Again, across all these other analyses, there are no statistically significant effects of HEN alignment on different Medicare expenditures. Finally, the Evaluation Contractor found no differences in the effects for different types of hospitals; which are shown in Appendix D.

6-13 For example, the difference-in-differences regression adjusted estimate for total expenditures (90 days) is $47.99, with a SE of $123.44. The mean total expenditures for an index discharge at a HEN-aligned hospital were $20,899.09 in 2013. With this SE, on the difference-in-differences impact estimate, the difference in total expenditures would not be significant at the 95 percent confidence level unless it was at least (123.44*1.96) = $241.94 in absolute value, which is just over 1 percent of total expenditures.

Table 6-6—Mean Medicare Expenditures per Discharge and Differences, by Expenditure Type and 90-Day Lookout Period, 2011 and 2013; n = 24,253,498

Expenditure Type	Unadjusted							Regression-Adjusted
	HEN			Non-HEN				
	2011	2013	Difference (SE)	2011	2013	Difference (SE)	Difference-in-Differences (SE)	Difference-in-Differences Impact Estimate (SE)
Total Expenditures	20,600.20	20,899.09	298.90** (41.25)	21,014.08	21,403.88	389.80 (221.52)	-90.90 (225.33)	47.99 (123.44)
Index Discharges	9,141.13	9,544.89	403.76** (26.58)	9,316.50	9,860.45	543.95** (115.08)	-140.19 (118.11)	-30.14 (77.58)
Post-Discharge	11,459.07	11,354.21	-104.86** (21.08)	11,735.61	11,522.79	-212.81 (142.50)	107.95 (144.05)	130.95 (87.31)
Post-Discharge IP	5,489.32	5,437.31	-52.01** (14.63)	5,529.53	5,458.42	-71.11 (69.94)	19.10 (71.45)	44.47 (51.98)
Outpatient	1,234.82	1,341.54	106.71** (4.03)	1,269.23	1,396.96	127.73** (17.92)	-21.02 (18.37)	3.86 (13.41)
DME	234.72	204.45	-30.27** (1.16)	238.64	206.78	-31.86** (2.93)	1.59 (3.15)	4.89 (3.30)
Home Health	991.65	983.89	-7.76* (3.56)	984.29	980.99	-3.29 (16.21)	-4.47 (16.59)	1.02 (15.77)
SNF	3,232.77	3,113.61	-119.15** (10.63)	3,531.31	3,293.21	-238.10* (102.31)	118.95 (102.86)	66.08 (54.45)
Hospice	175.38	175.97	0.59 (1.24)	186.44	182.63	-3.81 (7.03)	4.39 (7.14)	6.96 (8.14)
Professionals (i.e., Carrier)	100.41	97.43	-2.97** (0.23)	97.67	96.62	-1.05 (1.32)	-1.92 (1.34)	-1.58 (1.30)

Source: The Evaluation Contractor's analyses of Medicare claims data.

Notes: Each row corresponds to a separate difference-in-differences regression model using index discharge-level data, with discharges weighted using propensity-score based weights. Light blue expenditure types can be summed to obtain total expenditures. Orange expenditure types can be summed to obtain post-discharge expenditures. For all 10 outcomes, expenditures are price-adjusted and are expressed in January 2010 dollars. The first six columns of estimates present mean expenditures in 2011 and 2013, as well as the change from 2011 to 2013 for HEN-aligned and comparison hospitals using raw, or unadjusted, expenditures. The next column provides the difference in the change from 2011 to 2013 between HEN-aligned and comparison hospitals using raw, unadjusted, expenditures. The top number in the final column presents the main impact estimate, adjusting for patient characteristics, patient risk factors, and hospital fixed effects. The coefficient shows the average effect of HEN alignment in the 2013 period. The main impact estimates were calculated by using linear regression models. Appendix D provides the full list of controls included in the regression. Robust SEs, clustered by hospital, are provided in parentheses.

*Difference-in-differences treatment-comparison impact estimate significantly different from zero at the 0.05 level, two-tailed test.

**Difference-in-differences treatment-comparison impact estimate significantly different from zero at the 0.01 level, two-tailed test.

Table 6-7—Mean Medicare Expenditures per Discharge and Differences, by Expenditure Type and 180-Day Lookout Period, 2011 and 2013, n = 19,441,206

Expenditure Type	Unadjusted							Regression-Adjusted
	HEN			Non-HEN			Difference-in-Differences (SE)	Difference-in-Differences Impact Estimate (SE)
	2011	2013	Difference (SE)	2011	2013	Difference (SE)		
Total Expenditures	24,542.55	24,528.03	-14.52 (47.86)	25,067.45	25,192.16	124.71 (254.93)	-139.23 (259.38)	9.60 (175.19)
Index Discharges	9,102.89	9,383.65	280.76** (25.71)	9,272.17	9,720.60	448.42** (125.55)	-167.66 (128.15)	-46.43 (93.40)
Post-Discharge	15,439.67	15,144.38	-295.28** (30.01)	15,996.74	15,594.91	-401.83* (184.56)	106.55 (186.98)	122.83 (116.53)
Post-Discharge IP	7,403.24	7,241.61	-161.63** (20.31)	7,392.86	7,317.43	-75.43 (113.66)	-86.20 (115.46)	-46.06 (96.18)
Outpatient	2,118.55	2,264.48	145.93** (6.64)	2,186.33	2,348.30	161.97** (24.17)	-16.04 (25.07)	20.47 (21.28)
DME	406.54	350.87	-55.67** (2.01)	413.39	358.14	-55.25** (8.84)	-0.42 (9.07)	3.62 (9.34)
Home Health	1,405.58	1,375.48	-30.10** (4.96)	1,421.16	1,404.88	-16.28 (22.14)	-13.81 (22.69)	-6.66 (22.48)
SNF	3,719.66	3,531.45	-188.21** (12.22)	4,083.52	3,750.62	-332.90* (131.41)	144.69 (131.98)	60.83 (56.94)
Hospice	283.92	281.05	-2.87 (2.27)	304.95	288.23	-16.72 (16.36)	13.85 (16.52)	18.72 (18.37)
Professionals (i.e., Carrier)	102.17	99.44	-2.73** (0.25)	99.55	97.91	-1.64 (1.22)	-1.09 (1.24)	-0.83 (1.23)

Source: The Evaluation Contractor's analyses of Medicare claims data.

Notes: Each row corresponds to a separate difference-in-differences regression model using index discharge-level data, with discharges weighted using propensity-score based weights. Light blue expenditure types can be summed to obtain total expenditures. Orange expenditure types can be summed to obtain post-discharge expenditures. For all 10 outcomes, expenditures are price-adjusted and are expressed in January 2010 dollars. The first six columns of estimates present mean expenditures in 2011 and 2013, as well as the change from 2011 to 2013 for HEN-aligned and comparison hospitals using raw, or unadjusted, expenditures. The next column provides the difference in the change from 2011 to 2013 between HEN-aligned and comparison hospitals using raw, unadjusted, expenditures. The top number in the final column presents the main impact estimate, adjusting for patient characteristics, patient risk factors, and hospital fixed effects. The coefficient shows the average effect of HEN alignment in the 2013 period. The main impact estimates were calculated by using linear regression models. Appendix D provides the full list of controls included in the regression. Robust SEs, clustered by hospital, are provided in parentheses.

*Difference-in-differences treatment-comparison impact estimate significantly different from zero at the 0.05 level, two-tailed test.

**Difference-in-differences treatment-comparison impact estimate significantly different from zero at the 0.01 level, two-tailed test.

Discussion

The cost estimates of harm reduction at the national level, with the most conservative estimate totaling at least $8.67 billion, leave no doubt that the harm reductions described in Chapter 2 have benefitted the country financially as well as avoided individual suffering. Further, this is an underestimate, because most data from 2014 were unable to be included in time for this report; CMS intends to contract for additional evaluation to include data for the full PfP time period.

The difference-in-differences analyses in this chapter found no evidence that the HEN component of PfP affected various types of Medicare expenditures; that is, change in Medicare expenditures was similar between HEN and non-HEN-aligned hospitals including total expenditures as well as expenditures during the initial index admission or in the post-discharge period. These results are robust to the baseline year used; the type of expenditures examined; whether or not payments were price-adjusted; and the categorization of the PfP intervention period. There is also little evidence to suggest that certain types of hospitals were differentially affected by HEN alignment. However, 2014 data were unavailable in time for inclusion in the analysis; it is possible that hospitals and HENs became more effective over time such that results could be different if data for the full time period were able to be included.

One important caveat of the analysis is that only Medicare FFS expenditures are being examined. While no effects were found on these specific types of expenditures, it is possible that out-of-pocket payments, other payer expenditures, or actual costs of care were all affected. To the extent that Medicare FFS expenditures are correlated with these other measures, then the analyses could shed light on their impacts. However, future research looking at these other outcomes would be warranted to better understand if the HEN component affected these other costs and payments.

The approach taken is quite comprehensive in that it includes all individuals with an acute hospital admission and examines many diverse expenditures, which consequently allowed the capture of effects from HEN alignment on non-hospital expenditures and on types of patients not explicitly targeted. However, it is important to note, that a drawback of this approach is that it does not focus solely on those patients who had suffered or were at high risk for an adverse event, which could dilute the analyses. At the same time, the results presented in Chapter 4 did not find evidence that HEN alignment had an effect on different types of adverse events and harms.

Finally, as in all difference-in-differences analyses, one concern is that other hospital initiatives implemented in the time period of analysis, as well as broader healthcare policies and other economic elements, may all play a role in determining healthcare expenditures. If these other factors differentially affect HEN-aligned hospitals versus the comparison group of non-HEN-aligned hospitals in the sample period, this could mask the impact of HEN alignment and confound the difference-in-differences estimates.

7. Discussion

The preceding chapters (in parenthesis) presented many analyses based on a variety of data sources, aimed at answering these primary research questions.

- To what extent did inpatient harms and readmissions decrease nationally during 2012-2014, consistent with the Partnership for Patients (PfP) goal? (Chapter 2)
- To what extent did PfP play a role in the decrease? (Chapters 3, 4, 5, and 6)

The chapter synthesizes the results across the chapters to respond directly to these questions.

To what extent did inpatient harm and readmissions decrease nationally during 2012-2014, consistent with the PfP goal?

National rates of inpatient harm decreased. Rates of inpatient harm and Medicare fee-for-service (FFS) readmissions have markedly improved since the start of the PfP campaign. Inpatient harm shows reductions consistent with the PfP goal of reducing preventable inpatient harm by 40 percent by the end of 2014, while declines in readmissions fell short of the 20 percent reduction goal.[7-1,7-2] For many harm areas, improvement began prior to PfP. This is not surprising, since those who designed PfP selected harm areas to focus on for which there was already evidence that improvement was feasible, and other efforts to reduce these harms were underway. Whether the pace of improvement accelerated, remained steady, or slowed during the PfP campaign depends on the data source and measure used, although for many metrics and sources, the pace remained steady.

These harm reductions are estimated to have saved more than $8.67 billion. Two very different sets of estimates of the cost savings associated with the reductions in adverse events described above are available from two different sources. Each set of estimates was derived using different methods and data sources and included different adverse events and PfP focus areas. One set was developed by the Evaluation Contractor and the other by the Agency for Healthcare Research and Quality (AHRQ) (Chapter 6). Due to data availability issues, neither of the estimates covers all the PfP focus areas and neither covers the full PfP time period for all the harm areas. Future evaluation reports will include updated estimates based on the entire PfP campaign period of performance.[7-3]

The Evaluation Contractor has estimated that harm reductions nationally have resulted in at least $8.67 billion in cost savings between 2010 (the year before PfP began) through either 2013 or mid-to-late 2014 (varies by measure). This estimate is conservative. Many harms that were known to occur were not included in the cost savings estimate because there were no cost-per-event data from the literature that were well-matched to the type of harm being counted, so that the exclusions included adverse drug events (ADEs), stage 1 pressure ulcers (more severe pressure ulcers were counted), falls with injury other than fractures and harms that did not fall into the PfP focus areas. The estimate from AHRQ (which is less conservative) suggests savings may total $11.98 billion through 2013. Chapter 6 describes the methods behind the two

7-1 This goal is operationalized to be 17.6 percent reduction in the rate of adverse events, using an AHRQ estimate that 44 percent of inpatient harm may be preventable, and 40 percent of that equals 17.6 percent.

7-2 Medicare FFS readmissions dropped 5.6 percent through 2014 (see Chapter 2); non-Medicare data were available only through 2013 and show no decrease to that point (data provided by AHRQ).

7-3 Agency for Healthcare and Research Quality (AHRQ) Saving Lives and Saving Money: Hospital-Acquired Conditions Update Interim Data from National Efforts to Make Care Safer, 2010-2014. Rockville, MD; December 2015. Available at http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2014.html

estimates and shows the estimates for each harm area. It is unclear what proportion of savings in either set of estimates is attributable to PfP.

To what extent did PfP play a role in the decrease?

Environment of concurrent activity toward harm reduction and national scale complicated attribution of the decreased harm to any one initiative. The substantial improvements to patient safety occurred at a time when a number of initiatives to address patient safety problems were underway. The PfP campaign recognized most of these initiatives as aligned with its work including Quality Improvement Organization (QIO) initiatives, payment reforms, state-level initiatives, and private-sector initiatives. Many of these initiatives were likely aided by the spread of information fostered by PfP and vice versa. PfP actively coordinated with and encouraged these other efforts as partners in harm reduction. This active work with partners, along with the fact that the work toward similar goals was simultaneous, makes it very difficult to determine the contribution of any particular initiative to the national decrease in harm. At the same time, the national scale of PfP left little room in which to identify a comparison group of hospitals—a "counterfactual" against which outcomes from PfP hospitals can be compared, and an estimate of the campaign impact may be drawn.

PfP successfully reached hospitals with information on improving patient safety, and a majority of those aligned with a Hospital Engagement Network (HEN) took action intended to improve patient safety to some extent due to their participation. There was a high rate of hospital participation through aligning with (i.e., committing to work with) a HEN. A majority of the surveyed hospitals that aligned with a HEN considered themselves to be engaged in the work with the HEN, and a large number of hospitals nationally were very likely to report that they made changes to improve patient safety and readmissions as a result of their PfP engagement, citing it as more influential than other factors. The evidence is consistent with PfP having reached hospitals, through HENs, with information that was used by a large number of hospitals. From the perspective of many of the surveyed hospitals', the HENs were an effective approach for facilitating the spread of best practices and providing resources on a large scale. However, the ability to use the information effectively varied from hospital to hospital. At the national level, the Evaluation Contractor found no direct evidence that the HEN activities were linked to improvements with outcomes.

Evidence of PfP's impact on outcomes and costs is inconclusive. The Evaluation Contractor compared HEN-aligned and non-HEN-aligned hospitals, using a number of different data sources. Across the analyses, there was some evidence that HEN-aligned hospitals experienced changes associated with reduced patient harms during the timeframe of the PfP campaign. However, the results do not support the conclusion that the PfP campaign provided an independent impact to reduce patient harms in the targeted hospitals. This is not to say that the campaign did not work for any hospital; rather, the Evaluation Contractor found no evidence that the campaign improved key outcomes for the typical hospital. Table 7-1 summarizes the results for each harm area and aggregate measures of harm based on all data sets.

Harm area by harm area, the decreases listed in column 1 and discussed in detail in Chapter 2 mean that available all-payer measures show reductions consistent with achieving the PfP goal by the end of 2014 for ADE, venous thromboembolism (VTE), ventilator-associated pneumonia/ventilator-associated event (VAP/VAE), central line-associated blood stream infection (CLABSI), and obstetrical early elective deliveries (OB-EED), with Medicaid measures for catheter-associated urinary tract infection (CAUTI), all-

payer measures for surgical site infection (SSI), and all-payer and Medicare measures for pressure ulcers also showing progress consistent with achieving the goal.[7-4]

Whether the pace of improvement accelerated, remained steady, or slowed during the PfP depends on the data sources and measures used, although for many metrics and sources, the pace remained steady (right-hand column). For the measures exhibiting no change in trend, the results imply that pre-existing trends in the patient harm rates continued after the PfP campaign began. For measures in which trend improved or worsened, the change in these measures coincides with the beginning of the PfP campaign.

Table 7-1—National Trend Results by PfP Harm Area

PfP Harm Area	National Trend in Rate of Harm[c] (Data Source Abbreviation)	ITS Analysis of National Trend (Change in *Trend–Q1 2011 Breakpoint*)
ADE	**Decrease** AHRQ	**Improved Trend** AHRQ
VTE	**Decrease** Medicare (Claims), AHRQ, HENs (Self-Reported) **No Change** Medicaid (Claims, 2012)	**No Change** Medicaid (Claims)
Pressure Ulcers	**Decrease** NDNQI/CALNOC, Medicare (Claims), AHRQ **Increase** Medicaid (Claims, 2012)	**Worsening Trend** Medicaid (Claims)
Readmissions	**Decrease** Medicare (Claims), Medicaid (Claims, Q1 2013), AHRQ-Readm, HENs (Self-Reported)	**No Change** Medicaid (Claims, Children Q1 2013) **Improved Trend** Medicare (Claims), Medicaid (Claims, Adult 2012)
CLABSI	**Decrease** NHSN (Self-Reported Using CDC Case Definition), AHRQ, NDNQI, Medicaid (Claims, 2012) **No Change** Medicaid (Claims, Q1 2013)	**Improved Trend** Medicaid (Claims, 2012, Q1 2013)
CAUTI	**Decrease** AHRQ, NDNQI, Medicaid (Claims, 2012[a]) **No Change** NHSN (Self-Reported Using CDC Case Definition) **Increase** Medicaid (Claims, 2012[a])	**No Change** Medicaid (Claims, 2012, Q1 2013[b]) **Worsening Trend** Medicaid (Claims, Q1 2013[b])

7-4 The PfP goal is 17.6 percent reduction in harms (40 percent reduction of preventable harms, with 44 percent assumed preventable based on AHRQ's review of the literature pre-PfP), and there are roughly 4 years between the start of PfP and the end of 2014.

Table 7-1—National Trend Results by PfP Harm Area		
PfP Harm Area	**National Trend in Rate of Harm[c] (Data Source Abbreviation)**	**ITS Analysis of National Trend (Change in *Trend–Q1 2011 Breakpoint*)**
VAP	**Decrease** NDNQI, AHRQ	—
SSI	**Decrease** AHRQ, NHSN (Self-Reported Using CDC Case Definition)	—
OB-EED	**Decrease** NVSS, HENs (Self-Reported)	**No Change** NVSS
OB-Other	*OB Trauma* **Decrease** HENs (Self-Reported), Medicaid (Claims, 2012), AHRQ *Injury to Neonate* **No Change** HENs (Self-Reported) **Decrease** Medicaid (Claims, Q1 2013) *Assisted Ventilation* **Decrease** NVSS *APGAR Score 0-6* **No Change** NVSS *Admissions to NICU* **Increase** NVSS *Low Birth Weight* **No Change** NVSS	*OB Trauma* **No Change** Medicaid (Claims, 2012) *Injury to Neonate* **No Change** Medicaid (Claims, Q1 2013) *Assisted Ventilation* **No Change** NVSS *APGAR Score 0-6* **Worsening Trend** NVSS *Admissions to NICU* **Worsening Trend** NVSS *Low Birth Weight* **No Change** NVSS
Falls	**Decrease** AHRQ, NDNQI/CALNOC, Medicare (Claims)	**No Change** NDNQI/CALNOC, Medicare (Claims)
Aggregate Measures of Harm	**Decrease** AHRQ	**No Change** AHRQ

Table 7-1—National Trend Results by PfP Harm Area		
PfP Harm Area	**National Trend in Rate of Harm[c] (Data Source Abbreviation)**	**ITS Analysis of National Trend (Change in *Trend–Q1 2011 Breakpoint*)**

Source: Evaluation Contractor's analysis of data sources as listed.
Note: Data run through at least 2013 unless another end period is noted.
[a]In the Medicaid claims data, two measures were tested: the CAUTI HAC measure and a measure for Hospital-acquired UTIs (HAUTI)—which includes many more infections some of which are not related to catheters. For HAUTI, the pediatric rate decreased while the adult rate increased through 2012. For the CAUTI HAC, the reverse was true.
[b]In the Medicaid claims data, ITS analysis of change in trend using Q1 2011 as the structural break point showed no change for CAUTI –adults and HAUTI-pediatrics, and worsening for HAUTI-adults.
[c]"Decrease," "No Change," or "Increase" in the first column means a decrease was found through consideration of the preponderance of available evidence from ITS analyses, SPC chart evidence of special cause variation, or the visual trend showed the indicated direction.
Data source abbreviations (See Appendix B for a description of each source):
AHRQ = AHRQ National Scorecard, composed primarily of data from MPSMS obtained through chart abstraction.
AHRQ-Readm = All-payer readmissions rates calculated by AHRQ, provided through Noel Eldridge, personal communication, August 26, 2015.
Medicare = Medicare FFS claims.
Medicaid = Medicaid claims for 17 states.
NHSN = National Healthcare Safety Network, the Centers for Disease Control and Prevention (CDC).
NDNQI = National Database of Nursing Quality Indicators
NDNQI/CALNOC = NDNQI, supplemented by data on the same measures for about 100 additional hospitals from CALNOC.
NVSS = National Vital Statistics System.
HENs = Outcomes data submitted in aggregate form by Hospital Engagement Networks.

Evidence of the impact of the HEN component on outcomes and costs is inconclusive. The Evaluation Contractor conducted three types of analyses, one that examined national data series for structural breaks coinciding with the inception of the HEN contracts, another that analyzed changes in HEN-supplied data over time, and a third that estimated the direct impact of the HEN component by comparing HEN-aligned and non-HEN-aligned hospitals, using a number of different sources of data and models. Across the three analyses, there was some evidence that HEN-aligned hospitals experienced changes associated with reduced patient harms during the timeframe of the PfP campaign. However, the results do not support the conclusion that the PfP campaign provided an independent impact to reduce patient harms in the targeted hospitals. Table 7-2 summarizes the quantitative findings for each harm area and for the aggregate measures of harm.

The ITS analysis of national database measures detected significant structural breaks associated with accelerating reductions in readmissions for both Medicare and Medicaid claims-based measures. In contrast, none of the HAC data series exhibited structural breaks toward greater improvements in Q1 2012, the first full quarter of the HEN operations. Two harm areas, pressure ulcers and CAUTI, exhibited structural breaks toward worsening rates in the national database measures. For all of the harm areas, the ITS analyses did not detect structural breaks coinciding with the beginning of the HEN contracts.

The ITS analysis of HEN-supplied data examined whether there was an overall reduction in harm and an increase in the rate of that reduction after the start of the HEN component. Few outcome measures showed this pattern, with readmissions and OB-EED showing a moderate proportion of measures meeting these criteria. However, this analysis was limited by the fact that the HEN-supplied data varied across HENs and across harms in the measures used, as well as in the duration, periodicity, completeness, and consistency of hospital reporting.

In one of the difference-in-differences analyses (Bayesian), which examined four measures from Medicare claims data, there was a moderate probability of a slight or small impact for two harms (VTE and catheter-related blood stream infection [CRBSI], respectively), a moderate probability of substantial impact for one harm (pressure ulcers), and no impact for readmissions.[7-5] However, for all sets of the Bayesian estimates, the uncertainty interval around the impact estimates includes zero, so it is possible that the observed improvements are due to chance variation rather than HEN alignment. Frequentist difference-in-differences analyses of OB-EEDs and birth outcomes and of an aggregate measure of harms also found no measurable positive HEN impact. Analysis of Medicare FFS expenditures during and after a hospitalization found no difference in the change in those costs over time for beneficiaries admitted to HEN-aligned versus non-HEN-aligned hospitals. These results cannot be interpreted as evidence that the HEN component of the campaign did not work for any of the aligned hospitals, rather that little evidence was found that the HENs improved key outcomes for the average aligned hospital relative to changes being observed in non-HEN-aligned hospitals during this time period. Ironically, the more that PfP and its partners were able to reach all hospitals in the U.S., the less likely this type of analysis is able to estimate a different reduction in harm for HEN-aligned hospitals relative to non-HEN-aligned hospitals.

The ITS analysis used HEN-supplied data to examine whether there was an overall reduction in harm and an increase in the rate of that reduction after the start of the HEN component. Few outcome measures showed this pattern, with readmissions and OB-EED showing a moderate proportion of measures meeting these criteria. However, this analysis was limited by the fact that the HEN-supplied data varied across HENs and across harms in the measures used, as well as in the duration, periodicity, completeness, and consistency of hospital reporting. Importantly, OB-EED measures submitted by the HENs exhibited substantial reductions over the course of the campaign, with numerous measures exhibiting structural breaks consistent with a floor effect: the flattening of the trend as the measure approaches the lower boundary of zero OB-EEDs. This result after the PfP campaign efforts through the Strong Start Initiative and requesting commitments to reduce OB-EEDs is a compelling finding with respect to the timing and magnitude of results, and the potential impact of the PfP campaign.

Table 7-2—Summary of Quantitative HEN Results by PfP Harm Area

PfP Harm Area	ITS Analysis of National Trend (Change in *Trend–Q1 2012 Breakpoint*)	Difference-in-Differences Analysis (HEN-Aligned versus Non-HEN-Aligned)		ITS Analysis–Percentage of HEN Outcomes Measures Showing Both Overall Improvement and Accelerated Improvement Trend Following Start of HEN Component
		Type of Model	Result of Analysis of HEN Impact–Differential Change in HEN-aligned versus Non-HEN-Aligned Hospitals (Data Source)[c]	
ADE	—	—	—	**17%**
VTE	**No Change**	**Bayesian**	**Moderate Probability of HEN Impact of 2-5%** Medicare	**8%**

7-5 "Moderate probability" is 60 to 80 percent.

Table 7-2—Summary of Quantitative HEN Results by PfP Harm Area				
PfP Harm Area	ITS Analysis of National Trend (Change in *Trend–Q1 2012 Breakpoint*)	Difference-in-Differences Analysis (HEN-Aligned versus Non-HEN-Aligned)		ITS Analysis–Percentage of HEN Outcomes Measures Showing Both Overall Improvement and Accelerated Improvement Trend Following Start of HEN Component
		Type of Model	Result of Analysis of HEN Impact–Differential Change in HEN-aligned versus Non-HEN-Aligned Hospitals (Data Source)[c]	
Pressure Ulcers	**No Change** Medicare, Medicaid, NDNQI **Worsening Trend** NDNQI	**Bayesian**	**Moderate Probability of HEN Impact of 25%** Medicare	**9%**
Readmissions	**Improved Trend** Medicare, Medicaid	**Bayesian**	**No Evidence for HEN Impact** Medicare	**21%**
CLABSI	**No Change**	**Bayesian**	**Moderate Probability of HEN Impact of 5-10%** Medicare	**12%**
CAUTI	**No Change** NHSN, Medicaid (2012[b]) **Worsening** NDNQI, Medicaid (2012[b])	—	—	**9%**
VAP	**No Change**	—	—	**17%**
SSI	**No Change** NHSN	—	—	**6%**
OB-EED	—	**Frequentist**	**No Evidence for HEN Impact** NVSS	**27%**
OB-Other	**No Change**	**Frequentist**	**No Evidence for HEN Impact** NVSS	**8%**
Falls	**No Change**	—	—	**11%**
Aggregate Measures of Harm	—	**Frequentist**	**No Evidence for HEN Impact on Total Medicare Expenditures** Medicare **No Evidence for Large (20% or more) HEN Impact on Aggregate Measure of Harm** MPSMS	—

Table 7-2—Summary of Quantitative HEN Results by PfP Harm Area				
PfP Harm Area	ITS Analysis of National Trend (Change in *Trend–Q1 2012 Breakpoint*)	Difference-in-Differences Analysis (HEN-Aligned versus Non-HEN-Aligned)		ITS Analysis–Percentage of HEN Outcomes Measures Showing Both Overall Improvement and Accelerated Improvement Trend Following Start of HEN Component
		Type of Model	Result of Analysis of HEN Impact–Differential Change in HEN-aligned versus Non-HEN-Aligned Hospitals (Data Source)[c]	

Source: Evaluation Contractor's analysis of data sources as listed.
Note: Medicare claims data severely undercounts the number of harms relative to chart review, for the Bayesian analyses of VTE, pressure ulcers, and CLABSI (see Appendix D). However, the Evaluation Contractor is unaware of any evidence that the undercount results in a differential shortcoming for the HEN-aligned versus the non-HEN-aligned hospitals, and the readmissions analysis, OB-EED, OB-Other, and total Medicare expenditures analysis does not have this limitation.
[a]In the Medicaid claims data, two measures were tested: the CAUTI HAC measure and a measure for Hospital-acquired UTIs (HAUTI)—which includes many more infections some of which are not related to catheters. For HAUTI, the pediatric rate decreased while the adult rate increased through 2012. For the CAUTI HAC, the reverse was true.
[b]In the Medicaid claims data, ITS analysis of change in trend showed no change for HAUTI-pediatrics, and worsening for HAUTI-adults.
[c]High probability is 85 percent or more; moderate is 60 to 85 percent; and low is less than 60 percent probability.
Data source abbreviations (See Appendix B for a description of each source):
Medicare = Medicare FFS claims.
MPSMS = Medicare Patient Safety Monitoring System (chart review data abstracted by AHRQ).
NVSS = National Vital Statistics System.
HENs = Outcomes data submitted in aggregate form by Hospital Engagement Networks.

Since harms were reduced nationally during this time period, and over 80 percent of admissions nationally occurred in HEN-aligned hospitals, clearly improved safety within the HEN-aligned hospitals was a major part of the decrease. Results of the difference-in-differences analyses suggest that safety improved at an equal rate in the non-aligned hospitals. In addition, the cluster analysis, mixed models, and dose response models detected little in the way of any relationship between HEN activity and outcomes. There are at least two potential conclusions one could draw from these analyses.

Potential Conclusion 1: *It is possible that the design and implementation of PfP, working with and alongside other stakeholders and payment incentives to improve patient safety and reduce readmissions across all hospitals, was successful in reducing harms across HEN-aligned and non-HEN-aligned hospitals alike.*

The analytic results do not show, but are consistent with, the possibility that PfP worked exactly as intended. That is, the vast majority of hospitals in the country—both aligned and not aligned with a HEN—received the support and encouragement they needed from PfP, with and alongside the other aligned stakeholders and payment incentives, to address their patient safety opportunities.[7-6] It is also possible that by influencing the actions of such a large number of HEN-aligned hospitals, that hospitals that were not aligned took similar actions so as not to fall behind in the marketplace.[7-7] If this were the case, analyses aimed at detecting a difference in harm reduction between HEN and non-HEN-aligned hospitals would miss the impact of the campaign. Similarly, simply looking at trends in national rates since 2010 to identify the impact of PfP would miss the impact of other, unrelated, factors that influenced these outcomes. However, for many of the

7-6 Shafer et al. (2008) provide an example of another large-scale health quality improvement program where the focus of intensive assistance was on a relatively smaller set of hospitals, but the positive outcomes appeared to spread to other hospitals more broadly, possibly due to the broader roles of key partner organizations.
7-7 Burns and Pauly (2002) conclude that "the history of the adoption of managerial innovations and new corporate forms in the hospital industry reveals the strong presence of local imitation and industrywide bandwagon effects." (pp. 134-135)

measures examined, the data series examined did not contain sufficient historical data to determine if there were significant structural breaks following the beginning of the PfP campaign. For the measures where such data existed, the results are mixed.

Potential Conclusion 2: *HEN-aligned hospitals received and acted on more information than non-HEN-aligned hospitals, but without measureable payoff in terms of reduced harm outcomes and readmissions during the PfP period.*

Under PfP, a great deal of information was shared with hospitals. This information fell into three categories:

- Evidence-based.
- Self-described experience of a hospital or HEN that appeared to be having success based on a decreasing trend of harm post-intervention.
- Anecdotal, based on networking with other hospitals, often facilitated by the HEN.

There is no record of the extent to which either the efforts of HEN-aligned hospitals or other hospitals' efforts to improve safety were based on these three categories of information. It remains possible that a great deal of work based on anecdotal content and content that was short of evidence-based practice was both needed and accomplished equally by HEN-aligned and non-HEN-aligned hospitals to achieve the national reductions in harm. It is also possible that HEN-aligned hospitals worked more than others on changes where the basis was short of evidence-based, and that this did not earn them any better results than non-HEN-aligned hospitals who may have invested less effort and resources on changes to reduce harms; or that hospitals were just less effective in implementing additional changes successfully; thus explaining the lack of relationship between HEN activities and change in outcomes.[7-8] Additionally, there is evidence of substantial variation in the performance levels exhibited by the HENs participating in the PfP campaign. Thus, the similarities demonstrated between HEN-aligned and non-HEN-aligned hospitals on average, may mask the possibility that the quality of implementation, and hence outcome performance, differed from HEN to HEN.

Conclusion

In sum, when the time came to triangulate the results across the various analyses conducted to assess the impact of PfP to address the second research question, the results pointed in different directions.

- Qualitative and survey data pointed in the direction of PfP likely having had an effect on awareness of how the targeted harms can be reduced and on how care was delivered in a substantial portion of U.S. hospitals. PfP was perceived as a major force in supporting change on a widespread basis during 2012-2014, the same period when harms decreased nationally.
- Outcomes data from many sources and across many types of analyses pointed to little or no reduction in harms that can be directly attributed to the HEN component of PfP. In most cases, there was an equal reduction in harms in comparable hospitals that did not align with a HEN, many of whom were exposed to some of the same information and other public and private incentives and programs focused on harms reduction as the HEN-aligned hospitals.

Two features of the PfP campaign, the national focus and the robust implementation of its multiple partnership strategies, both key to achieving national patient safety goals, coupled with a variety of data and

[7-8] Kaissi and Begun (2008) suggest that while imitation of exemplars (i.e., hospitals responding to the type of information described in the second bullet point above) can sometimes be efficient and beneficial, "managers should carefully take account of local conditions that may distinguish their own organizations from the exemplars." Although in many instances, HENs worked with hospitals to adapt interventions to their local environment, it is not clear how much overall success this activity produced.

implementation constraints, limit the ability to capture the impact of the PfP campaign on outcomes from the quantitative analyses. Integrating the qualitative and quantitative analyses suggests several possibilities to explain the inconclusive impact analysis results for the HEN component, which include:

1. The campaign worked but the success could not be detected due to an unmeasurable effect on non-HEN-aligned hospitals resulting from the PfP partnership activities.
2. The campaign did not work better for HEN-aligned hospitals because other hospitals were able to access similar resources without the HENs.
3. The campaign did not work better for HEN-aligned hospitals because the additional actions they took were ineffective—well-intentioned but ultimately unsuccessful in reducing harms and readmissions.

As discussed in the next chapter, future programs could better shape and track what was tried and learned locally so as to be able to prevent the third possibility in the future. To prevent the uncertainty associated with possibilities 1 and 2, the Centers for Medicare & Medicaid Services (CMS) would need to build into future efforts more consideration of how to make the program more evaluable, potentially considering staggered roll-out of evidence-based next steps interventions among the hospitals that would benefit. However, there may also be circumstances when a program is needed, and where successful implementation would be compromised if it were designed to allow a rigorous evaluation. In such cases, CMS could plan the program focused on improvement with formative feedback and monitoring but without impact evaluation—with a goal and mechanism to track and signal when the program is no longer needed. In an era of limited resources, dealing with evaluability issues up-front in part through compatible program design will allow CMS to more efficiently achieve its most important goals.

8. Building on the PfP Experience

The Centers for Medicare & Medicaid Services (CMS) can build on the Partnership for Patients (PfP) experience to strengthen support for quality and safety improvement in hospitals, whatever programmatic form that takes. This chapter is divided between a section discussing considerations for future programs to support strong results, and a section discussing considerations for design features of large, complex initiatives to allow for stronger evaluation than was possible with PfP. These considerations have not been subjected to a cost-benefit assessment, so it is possible that the resources to implement any of them could be better spent elsewhere; policymakers will need to make those judgments as they weigh competing needs.

Considerations for Future Programs

The text below provides information about the PfP experience that CMS may wish to use in building future programs, related to:

- Features of PfP that Hospital Engagement Networks (HENs) reported worked well
- HEN activities associated with hospital response
- Considerations for strengthening implementation
- Considerations for additional research to support stronger implementation

These observations are presented as considerations rather than recommendations, because they lack strong evidence. However, they do represent feedback from program participants and observations by the Evaluation Contractor, with corroborating evidence noted where available. In one instance, participants' positive feedback on a feature of PfP ("pushes" within the overall campaign) was tested through interrupted time series (ITS) analysis, which did not confirm that outcomes improved with the pushes, so caution is warranted when participant feedback alone is used.

Features of PfP that HENs Reported Worked Well

During confidential interviews between the Evaluation Contractor and HENs in summer 2014, the HENs provided feedback about specific features of PfP. They reported the following features worked well and as such, CMS may wish to consider them in shaping future initiatives (see Chapter 3 for more detail).[8-1] Although HEN interviews were the primary source, where other sources of input were available, those are noted.

- **Bold aims, with interim targets and assessments**: Most HENs reported PfP's bold aims were helpful to their progress as a source of motivation, creating urgency and focus. Most also reported PfP's interim targets and assessments allowed them to gauge their progress toward the overall goals, keeping hospitals on track in their efforts. A management tracking system using HENs' monthly assessments of each hospital's status towards improvement targets suggested improvement trends showed steady progress during the months prior to the two interim assessments followed by a flatter slope after their completion.
- **Alignment with payment incentives and public reporting**: Most HENs reported CMS payment incentives generated a sense of urgency and accelerated hospitals' adoption of safety improvements.

[8-1] Participant feedback on usefulness is a weak basis for program design decisions, but stronger evidence for or against these features is not available, except as noted.

Payment incentives and public reporting were reported by hospitals to be influences on their patient safety actions, and HENs worked to amplify their effects.[8-2] Potentially related, several statistical process control (SPC) charts for national leading indicators exhibited shifts in center lines shortly after payment incentives were implemented (Chapter 2). All of these measures exhibited strong trends of decline prior to implementation of the payment incentives. However, it is unknown what proportion of hospitals implemented improvements in patient safety in anticipation of payment incentives beginning.

- **State/local public and private partnerships as a strategy for spreading best practices**: Partnerships were widespread, most frequently including Quality Improvement Organizations (QIOs) and county or state health departments, and often built on prior work together on quality or safety topics. Partners played a variety of roles, including supporting educational sessions and conferences, providing on-site consultation to hospitals, helping HENs carry out dissemination campaigns, helping to reach key stakeholder groups through constituent members, and engaging with HENs on strategic program planning. In addition to highlighting partnerships as helpful to their progress in the summer 2014 interviews, many HENs reported, through a detailed listing of specific partnerships and the experience with them, that many of their partners' work was moderately or tightly integrated with their own, and classified many of the partnerships as "essential" to their progress.
- **Support contractor assistance:** Most HENs reported that the work of PfP support contractors (the National Content Developer [NCD], Patient and Family Engagement Contractor [PFEC], and the Evaluation Contractor) had helped their progress, for example, by connecting them to high-quality speakers and content, offering the opportunity for their staff and hospitals to attend an in-depth series of master classes on patient and family engagement (PFE), and providing data feedback reports to allow them to compare their progress to others.
- **CMS-focused "pushes" within a broader campaign**: Most HENs reported that the CMS pushes, whereby CMS asked HENs and hospitals to focus intensely on reducing a specific adverse event and report progress back to CMS on a short timeframe (such as 3 to 6 months), added urgency and resulted in more intensive programming for obstetrical early elective delivery (OB-EED), readmissions, and catheter-associated urinary tract infection (CAUTI). However, ITS analysis conducted by the Evaluation Contractor using the HENs' data did not find evidence that the pace of improvement on outcomes accelerated after the pushes.

HEN Activities Associated with Hospital-Reported Response

Survey data from HEN-aligned hospitals suggest several common types of HEN activities were helpful in generating hospital responses. More specifically, hospitals that received each of the following HEN services were more likely than those that did not receive these services to report making operational changes to reduce harm or readmissions due to PfP (Survey of Participation in Patient Safety Activities, see Chapter 3 for more detail). While this approach is biased towards overstatement of the effect of these activities, the differences are large enough that CMS may wish to consider this feedback in considering future supportive strategies.

- **Peer-to-peer networking**: HENs' facilitation of networking was strongly associated with hospitals making changes due to participation in PfP. On average, hospitals participating in value-added networking were about three times as likely to report making changes due to PfP in the area they networked as HEN-aligned hospitals that did not participate in the networking activity. In addition, on

[8-2] Results from the Survey on Prevention of Adverse Events and Reduction of Readmissions, reported in Chapter 3.

open space for comments in the survey, 30 percent of the hospitals offered positive comments (nearly 300 hospitals) that complimented the collaboration/networking opportunities.[8-3]

- **Data feedback**: On average, hospitals receiving data feedback from their HEN were about three times as likely to report making changes due to PfP in the areas with the feedback as other HEN-aligned hospital that did not receive the feedback.
- **Skills training**: On average, hospitals receiving skills training through their HEN were more than twice as likely to report making changes due to PfP in the area where they received the training as HEN-aligned hospitals that did not receive skills training for that area.[8-4]
- **Virtual consultation/coaching**: On average, hospitals receiving virtual consultation/coaching were more than twice as likely to report making changes due to PfP in the relevant area as those that did not receive consultation for that area. Also, during discussions in fall 2014 about their interventions, 13 HENs expressed that a lesson learned had been the importance of individualized assistance to hospitals (more may have agreed this was important but did not specifically express it in the interviews).
- **Provision of educational resources and tools**: Hospitals receiving educational resources/tools were more than two times as likely to report making changes due to PfP in the relevant area as those who did not receive consultation for that area. Also, on open space for comments, 46 percent of those offering positive comments about PfP (nearly 450 hospitals) commended PfP's dissemination of resources and information, including toolkits, webinars, and websites/repositories.
- **On-site visits**: Hospitals receiving on-site visits were nearly two times as likely to report making changes due to PfP in the relevant area as those that did not receive consultation for that area.

Considerations for Strengthening Implementation

Based on participant feedback and observation by the Evaluation Contractor, CMS should consider refinements or new approaches aimed at strengthening implementation by increasing health information technology (IT) support and enhancing the business case for harm and readmissions reduction.

Health IT Support

Over one-fourth of hospital survey respondents (28 percent) identified a need for better health IT to further improve patient safety and reduce readmissions, consistent with the finding by the ECRI Institute that health IT is a top-ten patient safety concern.[8-5,8-6] Both HENs and visited hospitals pointed to how established electronic health records (EHRs) can facilitate change by supporting measurement and standardizing implementation of interventions. However, in the short-run, HENs cited hospital EHR implementations as diverting attention from harm reduction work and disrupting data reporting, and several visited hospitals provided illustrations of how temporary patient safety problems can occur following health IT system changes. Future CMS quality enhancement initiatives can support the potential of health IT through continuing intra-agency collaboration and communication (e.g., working with Office of the National Coordinator for Health Information Technology [ONC] and the Agency for Healthcare Research and Quality

8-3 For context, 2,432 hospitals participated in the survey, and 40 percent commented. Of the comments, 86 percent were positive, 5 percent negative, and 9 percent mixed (containing both positive and negative aspects of the comment).

8-4 This is the simple average of the odds ratios for skills training across all 11 adverse event areas (see Chapter 3).

8-5 The survey question was open-ended, allowing hospitals up to three responses: "What tools and resources is your hospital lacking that would be helpful in pursuing reductions in adverse events and readmissions?"

8-6 The top-ten list is compiled based on the staff's experience investigating device-related incidents, evaluating medical devices in the ECRI Institute's testing laboratory, and reviewing reports from ECRI Institute's and other organizations' databases for medical device problems and patient safety events. (ECRI 2015)

[AHRQ]) regarding the monitoring, development, and research around the ways health IT can help support patient safety improvement and avoid harm.[8-7]

Work to Enhance the Business Case

The fact that HENs identified resource constraints within their hospitals as a barrier to achieving all that they otherwise might have been able to achieve suggests a remaining need to enhance the business case for hospitals to invest in harm and readmissions reduction. In addition, HENs and state hospital associations (SHAs) often mentioned the economic challenges, particularly to small hospitals, from readmissions reduction due to reduced admissions. It will be important to carefully weigh the balance between continued monitoring and refinement of incentives so as to provide the necessary incentives without endangering the economic condition of hospitals aiming to "do the right thing," as CMS continues with these strategies.

Considerations for Additional Research to Support Future Patient Safety Programs

Additional research may be useful to address remaining needs around the evidence base for certain patient safety areas, as well as to follow up on the disappointing readmissions analyses results. The research to enhance the evidence could be included as part of a broader program of work, and could incorporate randomized design features and/or staggered roll-out of the interventions being tested, to ensure that the interventions' effects on outcomes can become known. The following two sections provide considerations for improving this process.

Undertake Additional, Smaller-Scale Research to Improve the Evidence Base for Interventions Regarding PFE and Adverse Drug Events (ADE)

CMS should consider smaller-scale research to further develop the evidence about what interventions are effective for PFE and ADEs prior to additional national-level intervention on these topics. Many other areas of patient safety would also benefit from additional research, per AHRQ's summary of the evidence base (AHRQ 2013). Enhancing the evidence in these two areas were viewed as particularly relevant follow-ups to the PfP experience, because of the strong cross-cutting focus on PFE under PfP, and HENs' comments regarding their struggles with how to best address ADEs along with its continued importance as a major source of inpatient harm.

Although PFE is believed to be a potentially powerful tool in harm prevention, and some progress was clearly made under PfP in involving patients and families, the evidence base for what strategies for involvement are most effective and how to measure results are not yet known (AHRQ 2013).

Despite a substantial drop in ADEs nationally during the PfP period, ADEs still represented one-third of all inpatient harm measured in 2013 in AHRQ's National Scorecard, occurring in an estimated 4 percent of all discharges. This level of harm is greater than any other single measured area of harm. Measurement issues regarding ADEs were cited by many HENs as a barrier to effective intervention in this area; most HENs therefore focused on measurement as a starting point, and encouraged hospitals to involve pharmacists and conduct case-by-case follow-up to instances where the measurement identified potential problems. Given the HENs' experiences in this area, it seems likely that the drop did not represent organized, evidence-based intervention, but rather progress from the new focus on this topic in a general way (such as through greater

8-7 One such effort, for example, is the ONC's Health IT Safety Center Roadmap, released July 2015 (ONC 2015).

general awareness among clinicians), greater use of newer drugs with less potential for adverse effect, the increased implementation medication therapy management (MTM) programs, or some other factor. Additional gains may require interventions not yet defined.

Undertake Additional Readmissions Research to Clarify Data Patterns and Necessary Interventions

CMS should consider supporting additional readmissions research on two fronts:

1. Clarifying the complexities of readmissions trends and patterns that underlie the lackluster drop in readmissions despite a robust effort to reduce them, and
2. Clarifying the essential pathway to reducing readmissions, given the complexity of the full set of recommended interventions.

Clarifying Readmissions Trends and Patterns

Although the clear downward shift in Medicare fee-for-service (FFS) readmissions and smaller drop in all-payer readmissions is a good start, the lack of ability to come anywhere near the ambitious 20 percent reduction goal on this prevalent and costly outcome suggests follow-up research to examine the causes. Beyond the national trend, HENs and HEN-aligned hospitals were very heavily engaged in readmissions reduction activities, so it is surprising to find they reduced Medicare readmissions less (very slightly less) than non-aligned hospitals, although given the very small magnitude of the finding, future research should probably focus on the larger question of why both sets did not come close to goal.[8-8] At least two branches of potentially useful research could be pursued; additional ideas may emerge during discussion of these results.

Further Analyses to Uncover Reasons for the Gap Between Goal and Observed Trend

Reasons that could be analyzed include:

- *The measure definition*: the measure used a traditional denominator of index admissions for the readmissions rate; the HENs encouraged their hospitals to work with long-term care settings as well as other post-acute providers in ways that could have reduced admissions (the denominator) as well as readmissions. Health care improvement efforts more broadly are likely more important influences on admissions rates, but the HENs could have played a role. Broader measures of readmissions per 1,000 beneficiaries could be explored and be presented with the current measure to provide additional context. Measures could also be used which incorporate emergency room and observation days, to explore arguments that reduction is due to switching from admission to emergency room or observation stay care (Himmelstein and Woolhandler 2015), although Daughtridge, Archibald and Conway 2014 provide data that suggest this does not account for all the Medicare FFS readmissions reduction.
- *The relationship of readmissions to patient needs and community support factors*: these factors, which are mostly not within the hospital's control, may be stronger influences on readmissions than previously realized, and may have made it much tougher than anticipated for hospitals to reduce

[8-8] However, if CMS would like to further delve into possible reasons for the lack of demonstrated impact, several foci are possible: comparative difference in all-payer readmissions (did HENs reduce those, even if they did not have an impact on Medicare readmissions?); the role of QIOs, which could have played a greater role with non-HEN hospitals so as to avoid duplication of effort; and the possibility that HEN hospitals' focus on 30-day *all-payer* readmissions diluted the finite resources that were mostly targeted to prevent Medicare readmissions in non-HEN hospitals.

readmissions. Their role could be clarified, though this would require an investment in primary data collection as well as additional analysis of existing data.

- *The relationship of readmissions to end-of-life care and mortality*: There may be complex relationships between readmissions and mortality; for example, some research has detected an inverse relationship for heart failure patients, whereby fewer readmissions is associated with higher mortality (Gorodeski, Starling, and Blackstone 2010; Heidenreich et al. 2010), although this does not hold for all conditions (Tsai et al. 2013). Could hospitals be lengthening life for some patients as they focus on improving post-discharge care, although not reducing readmissions as much as originally thought to be optimal?
- *Further analyses to uncover and support essential pathways to reduced readmissions*: The number of recommended and implemented intervention steps for a hospital to reduce readmissions is very large, including risk assessment, care coordination, discharge planning, medication reconciliation, patient and family engagement, post-discharge follow-up, care transitions models, and work with community partners. Not surprisingly, hospitals with support of HENs commonly "picked and chose" what they could do with available resources. Across the HENs, 42 specific, intervention types aimed at readmissions reduction were used by at least 30 percent of the hospitals in at least on HEN (see Appendix E), with 17 different types reaching this level of take-up in at least 10 HENs. Yet the lack of stronger results suggests it may not have been enough, or enough of the right things. Greater clarity is needed about the essential set of interventions needed to produce results in hospitals with common profiles of readmissions or communities (since not every hospital will have the same issues).

Recommendations for Approaches to Evaluation of Large-Scale Quality Improvement Programs

Large-scale quality improvement initiatives such as PfP are best characterized as complex adaptive systems.[8-9] They are complex because they are not hierarchical, not governed by one decision-maker or according to one set of rules; instead, decisions are made by a large number of different stakeholders with different incentives, expertise, and expectations who employ different approaches and operate in different markets. Such systems are considered adaptive in part because individual and collective behaviors change naturally in unpredictable ways over time. Even when primary goals are well-defined, the specific pathways chosen to achieve the goals are likely to be refined and evolve over time.

Evaluation of such complex adaptive systems would be strengthened by greater attention to building systems for experimentation and data collection within the larger context of the initiative, while also maintaining the flexibility for innovation and adaptation to local circumstances. It is possible to design large-scale initiatives that encourage flexibility and creativity while maintaining evaluability of the program.

Building on the challenges faced in evaluating the PfP campaign, the Evaluation Contractor recommends several ways that CMS could develop a scalable and repeatable system for the evaluation of complex adaptive systems as it moves forward with developing future large-scale quality improvement projects. Specifically, the following components could be developed and planned prior to or in the early stages of implementation of the specific program:

- Data systems – sources and collection
- Data validation
- Metric development, evaluation, and implementation

[8-9] Rouse, WB. Health Care as a Complex Adaptive System: Implications for Design and Management. *The Bridge*. 2008; 38(1):17-25.

- Measurement strategy
- Documentation of learning and spread of information
- Systematic focus on micro-experiments, locally-designed actions, Plan-Do-Study-Act (PDSA) cycles
- Methods for improving the evaluator's ability to link processes to outcomes
- Pathways for identifying and disseminating best practices innovations
- Methods for tracking and evaluating leadership
- Methods for tracking evolution of collaborations and partnerships

Data Systems—Sources and Collection

Large scale quality improvement campaigns such as the PfP have the potential to generate large volumes of data collected locally at the provider level. Assembling and managing such data for the purpose of evaluation requires forethought of the types of data that will be provided, the nature of the interventions to be evaluated, and the analyses that will be performed during the evaluation task. CMS should consider specifying minimum acceptable data quality and submission standards to impose an acceptable degree of structure on the data being collected for evaluation purposes. While contractors could be allowed the opportunity to begin with flexible designs, CMS would ideally encourage or mandate that contractors actively establish a standardized data-reporting platform or template within a reasonable time frame. Contractors could have active input into this design so that it poses the least burden possible, while also meeting the needs of the evaluation.

Data Validation

The use of locally collected quality improvement data that is self-reported by participating clinical providers requires that evaluators work to verify the accuracy of the data reported. Therefore it would be beneficial for CMS to build a data auditing task into new projects. The audit should be designed and initiated early in the campaign timeline to allow for adjustments and corrections to be made that will improve the quality of data submitted in subsequent reporting periods. The audit protocol should aim to collect an adequate amount of data in the appropriate structure to provide sufficient statistical power for analyses, while minimizing the burden on data providers. Auditing processes might consist of randomly selecting providers and requesting access to electronic health records (EHR) for verification of measure rates, or using visual chart abstraction to validate rates derived from EHR or claims data. Furthermore, CMS should communicate the audit expectations to contractors during the solicitation phase of the program to maximize cooperation by participants and ensure that subcontract requirements for audits and data reporting are aligned.

Methods for Real-Time, or Near Real-Time Evaluation of Data

The ability to respond to changes in complex systems in an agile manner requires feedback, which requires data. In addition to suggesting methods of collecting data, and methods of designing measures, CMS should work with the evaluation contractor prior to campaign implementation to select the data sources and metrics that will be monitored on an ongoing bases. Statistical process control (SPC) charts have proved very useful in providing timely and relevant data analysis, and might be considered as a preferred method for monitoring and rapid feedback. Additionally, effective feedback loops will need to be designed so that campaign implementation issues may be identified early and resolved prior to becoming a detriment to campaign functioning. As campaigns evolve, the focus of monitoring activities will naturally change. CMS will need to work with evaluation contractors to plan and implement changes to monitoring systems in such a way that does not disrupt or interfere with processes already under surveillance.

Metric Development, Evaluation, and Implementation

In order to optimize flexibility at the hospital/provider level, while operating a program that can be rigorously evaluated, the Evaluation Contractor recommends requiring that participants clearly define self-reported process and outcome measure specifications, and evaluate their quality relative to other potential measures early on, to promote evolution toward standardization of measures early in the program. Minimum specification standards would including diagnostic and procedure codes and value sets used to calculate numerators and denominators, as well as clearly defined scaling units (e.g., per 1,000 discharges or per 10,000 patient days, etc.) and any risk adjustment algorithms used to define expected values for standardized infection ratios (SIRs) and observed-to-expected ratios. Minimum reporting standards should also be established and incentivized, encouraging contractors to obtain data on the same measures from the large majority, if not all, clinical providers. Additionally, contractors should be incentivized to promote consistent reporting over time by all providers. Allowing clinical providers to move into and out of the program or allowing them to stop reporting data should be discouraged.

Preliminary decisions should also be made regarding which, if any, measures are to be constructed by the Evaluation Contractor from external data sources. Data collected at the local level, by individual providers, will allow for more immediate feedback on campaign performance, but is susceptible to greater variability in measure specifications and data structures. National data sources have a greater time lag than locally-collected data might have, but have the advantage of comparability. However, national data sources are more likely to be collected and available in a standardized format, with clearly defined specifications. Still, national sources such as Medicare claims data, the Centers for Disease Control and Prevention's (CDC's) National Healthcare Safety Network (NHSN), and the National Database of Nursing Quality Indicators® (NDNQI®) data each report different measures and have specific strengths and weaknesses.[8-10,8-11]

Measurement Strategy

To the extent possible, future initiatives should provide a strong up-front map of desired measurement, that is, definitions of the desired measures and the timeframe for moving to them, even if it is not possible for all involved providers to produce the desired measures at the start. PfP offered near-complete flexibility in HENs' and hospitals' selection of outcome measures, and then encouraged movement toward common and better measures over time. Although many hospitals might not have participated had there not been flexibility in measurement at least at the start, many HENs perceived this encouragement toward common and better measures as a shift in policy that caused frustration and additional work. Creating an up-front measurement map or list may require contractor support and thus the relevant support contract may best begin prior to going live with the initiative. This would better support program execution through better benchmarking, as well as its evaluability.

Documentation of Learning and Spread of Information

Documentation of learning under PfP was primarily retrospective and incomplete, resulting in analytic challenges in tying interventions to results. For future initiatives, CMS should consider supporting central, systematic, concurrent exchange of information about take-up of evidence-based practices and other, novel or adapted interventions as well as what was learned by the hospitals implementing them. HENs focused on connecting their aligned hospitals with a multitude of resources, but—with rare exceptions—there was little

8-10 NDNQI® is a registered trademark of the American Nurses Association (ANA). NDNQI® data were supplied by ANA. The ANA disclaims responsibility for any analyses, interpretations, or conclusions.

8-11 See Appendix B for a detailed discussion of data sources used in this report. Other data sources may be preferable depending on the subject of the quality improvement campaign being evaluation.

focus during the campaign on documenting what hospitals did with those resources and what results they found, except for collecting examples of positive experience accompanied by a trend of reduced harm, where such examples could be located. The same is true with respect to documenting collaborations and partnerships with external influences, including recording basic information about their identity, their timing, intensity, and the nature of the collaborations. Building a more systematic, concurrent knowledge base covering both success and lack of success in implementing specific types of interventions should help both participants and CMS understand what is being learned and identify challenges and barriers to effecting changes.

Systematic Focus on Micro-Experiments, Locally-Designed Actions, PDSA Cycles

A key to strong learning within any quality or safety-focused learning community is the ability to effectively document and share experience with PDSA cycles, that is, improvement attempts and experiences that describe the goals, the intervention, refinements of the intervention based on rapid-cycle measurement, and subsequent results.[8-12] The Evaluation Contractor observed struggle within the PfP learning community to find and coach potential speakers on how to share their experiences. It is very likely that many opportunities were missed because the underlying content was lacking—that is, research has shown that many hospitals do not systematically follow the best practice steps for PDSA cycles or document their experiences well (Taylor et al. 2013). To strengthen sharing within the learning community in the future, as well as to enhance the ability of hospitals to conduct effective improvement cycles, CMS may wish to consider supporting training and tools to bring more hospitals and clinicians within hospitals up to a minimum level of competency with the PDSA process, creating a new norm for quality of sharing. Various tools and resources are available, ranging from simple tools such as the Institute for Healthcare Improvement's (IHI's) PDSA worksheet, to books (such as Provost and Murray 2011), to intensive training on Lean six sigma technique. HENs did facilitate this type of PDSA training to some extent, but more could be done to strengthen this aspect of implementation.

By focusing effort on developing localized expertise in quality improvement research, CMS would instill and promote a stronger culture of quality improvement across providers on a national scale, while simultaneously improving the quality of large-scale interventions relying heavily on local data. Importantly, the focus on experimentation is not intended to restrict front line staff from implementing innovative strategies, but to encourage them to report the information in a manner that can contribute to the overall evaluation effort.

Mechanisms for Improving the Evaluator's Ability to Link Processes to Outcomes

A clinical contractor can do more to support the meaningful evaluation of the success of its interventions than simply adopt and report a process and an outcome measure. Again, learning from some of the challenges faced in evaluating the PfP campaign, HENs and hospitals should be given the tools necessary to overcome the most frequent challenges to collecting data for a more useful evaluation. For example, where multiple processes are tried in a single hospital, but only one outcome measure is collected, attribution of quality improvements between the methods will be difficult at best, unless processes are altered with a temporal sequencing that allows examination of potential impacts net of other interventions. Similarly, if multiple outcomes are measured among different populations or time periods, then process measures for each population should be collected and reported in order to understand the changes in healthcare delivery for each population relative to changes in outcomes. Data permitting the Evaluation Contractor to discern which patients actually received an intervention and whether their outcome changed could be collected. Crossover-

[8-12] See, for example, Taylor MJ, McNicholas C, Nicolay C, et al. Systematic Review of the Application of Plan-Do-Study-Act Method to Improve Quality in Healthcare. BMJ Qual Saf 2013;0:1-9. Available at http://qualitysafety.bmj.com/content/early/2013/09/11/bmjqs-2013-001862.full.pdf%2Bhtml, accessed on August 26, 2015.

style designs in which HENs implement multiple strategies concurrently, but in different subgroups of hospitals provide a method for strengthening the links between process and outcome. Implementation of this type would allow a single HEN to try out numerous strategies in different focus areas, while maintaining a higher standard of rigor for identifying outcome changes due to replication. This type of crossover design also allows for immediate replication of results when successful interventions are rolled out to other hospital subgroups.

Careful attention to and documentation of changes made at individual hospitals at specific points in time in response to specific HEN activities is crucial for demonstrating a stronger temporal connection between when hospitals became aligned with HENs, the evidence-based patient safety processes that each hospital initiated, and the outcomes that each hospital achieved.

CMS should therefore provide a method for keeping track of what materials/tools/suggestions HENs send to hospitals, when, and how (or whether) they were implemented or spread further to other hospitals, in addition to whether there is an impact on patient outcomes. These should be recorded in a standardized format, even if the actions taken are not standardized.

Pathways for Identifying and Disseminating Best Practices/Innovations

One of the fundamental design principles identified for planning complex adaptive systems is to maximize strategies for allowing complexity where it can be managed best, while decreasing complexity for end users. This is especially pertinent to large scale quality improvement projects in health care because of the great complexity of information developed during these programs. Focusing on providing as many contributions to safety, and as many ideas as possible was one of the goals of the PfP campaign. The Evaluation Contractor recommends that planning for additional work at a higher level (analogous to a HEN or contractor level) in identifying the best solutions and presenting them to hospitals in the simplest way possible would greatly improve the value of the program in terms of the results to the end users, both the hospital personnel providing care, and the patients receiving it. Such pathways might make use of partnerships with external stakeholders, such as clinical experts in something analogous to a Technical Expert Panel; they might seek to engage patients and families in evaluating approaches most in line with patient-centered care.

These pathways need to be monitored to evaluate the effectiveness of dissemination pathways, collecting data on timing, and identifying responsible individuals, partners, or network connections. As previously described, these pathways should provide a method for keeping track of what materials/tools/suggestions clinical support contractors such as HENs send to providers, when, and how or whether they were implemented or spread further to other providers. Ultimately, the data could also be connected to observable changes in outcomes to more effectively assess which components of the campaign may have had an impact on patient safety.

Methods for Tracking Evolution of Collaborations and Partnerships

CMS in general, and the PfP campaign in particular, had an overarching goal of promoting collaboration and partnerships among hospitals, and HENs, as well as other CMS programs, and other stakeholders. The theory behind promoting such collaborations is that the intentional alignment of forces on specific problems would increase the likelihood of achieving successful quality improvements. To evaluate the value added by such collaborations and partnerships requires basic information about their identity, their timing, intensity, and the nature of the collaborations. For the PfP campaign, these data were collected in a retrospective manner, and may not reliably capture all of the relevant information for each contractor. The data would be far more accurate and comprehensive if collected on an ongoing basis throughout the campaign.

Methods for Tracking and Evaluating Leadership

Another key evaluation metric for complex adaptive systems is the identification of leaders.[8-13] Since no one entity is in charge of the complex interrelationships between the multiple participants in the health care system, it has been suggested that the management approach should emphasize leadership, the ability to inspire personal commitments and the development of close lateral relationships to improve efficiency and effectiveness. Identifying which agents of change the system recognizes as de facto leaders could provide insight into how best to provide incentives for future behavior or to continue changes made during the quality improvement program. Early identification of leaders may also allow campaign organizers to efficiently address bottlenecks and breakdowns in network functioning, and quickly work to replace key personnel when leadership changes occur.

Evaluation

Evaluation of this comprehensive system would examine the structure of the system, documenting its evolution over time, and the relationship between system activities and goals. This would include identifying and tracking the major components—the support contractors (e.g., HENs) and providers or other facilities. Their connections and interfaces would be evaluated, together with indicators such as the density, directionality and reciprocity of their communications. Information on the timing of collaborations and initiatives would also be collected, as well as the nature of collaborations, how they manifested in an improvement of treatment processes or outcomes, and whether value was added by the partnership.

Ideally, there would be ongoing evaluations and feedback regarding the functioning of each component of the system, specifically determining whether the support contractors are functioning as they were planned to function. This would answer questions such as are the support contractors actively involved with the providers? Are they providing education, tools, leadership, etc.? Are contractors making sense of feedback from the providers and contributing value in terms of simplifying the complex information and feedback flows?

On the other hand, are the clinical providers (e.g. physicians, hospitals or other providers) doing what they should? Are they implementing and testing strategies, providing data, and making use of feedback to adapt their processes?

Another element would be evaluation of the functioning of the interface between system components: how well do the components interact with each other? This might measure characteristics such as the directionality, reciprocity, and frequency of contact; responsiveness to issues presented; value added in terms of identifying and disseminating best practices; and value added in terms of testing creative initiatives and reliable data collection and reporting strategies.

Finally, the system must recognize that large scale quality improvement initiatives may be successful or not on a variety of levels and in a variety of ways. CMS may consider the overall campaign-level success in achieving goals. However, such campaign-level metrics will generally not be informative regarding the variation of success exhibited by specific providers in specific topic areas. Evaluating the implementation of such a campaign will require strategies to evaluate the campaign at multiple levels, using different approaches. These might include the following types of questions:

- Local implementation: Is it working? To what extent was hospital X able to reduce patient harms? Was intervention Y associated with reductions in patient harms in hospital X?

[8-13] Rouse, WB. Health Care as a Complex Adaptive System: Implications for Design and Management. *The Bridge*. 2008; 38(1):22-23.

- HEN/Mediation contractor functioning: Was the system able to simplify decision-making for end users (hospitals) and add value in terms of narrowing the search for best practices? Did the contractor identify best practices in a timely manner for spread across the hospitals? Did it provide feedback to hospitals to enable them to judge their success?
- Network functioning: Did the system support dissemination of materials in a timely and effective manner for spread of best practices?
- Campaign level: Net of other programs and campaigns occurring at the time, did the campaign generate greater improvements among aligned hospitals? Is there supporting dose response analysis at the campaign level?

Conclusion

A cornerstone of the approach to quality improvement used by the PfP campaign was the concept of rapid-cycle improvement, or the PDSA cycle.[8-14] The PDSA cycle uses the following steps: planning a change to improve a system, making incremental changes to improve existing processes, observing the results of those changes, and then using the results to plan the next target for improvement changes. This order of steps repeats over time in a continuous cycle of quality improvement effort.

The PfP campaign used this tool along with others to promote improvements in patient quality on a scale not previously attempted. The scale of the campaign was massive – over 3,700 hospitals, hundreds of federal, state, local, and private partners, and a scope that spanned 11 diverse patient harm areas. The implementation of such a large endeavor, however, did not come without its challenges and the PfP campaign adapted to those challenges on an ongoing basis. Drawing on the philosophy of Deming and Shewhart, this report represents the completion of the primary task in the study segment of the cycle. The recommendations and considerations discussed here are provided as input for the next segment of the cycle, in which future large scale quality improvement initiatives in health care are designed and implemented.

The analyses presented in this report represent the culmination of a diverse array of methods used to understand how the PfP campaign worked and the extent to which the campaign generated intended results. Triangulation of the findings across these diverse analyses provides two important results that point in different directions.

- Qualitative and survey data pointed in the direction of PfP likely having had an effect on awareness of how the targeted harms can be reduced and on how care was delivered in a substantial portion of U.S. hospitals. PfP was perceived as a major force in supporting change on a widespread basis during 2012-2014, the same period when harms decreased nationally.
- Outcomes data from many sources and across many types of analyses pointed to little or no reduction in harms that can be directly attributed to the HEN component of PfP. In most cases, there was an equal reduction in harms in comparable hospitals that did not align with a HEN, many of whom may have been exposed to some of the same information, other public and private incentives, and programs focused on harms reduction as the HEN-aligned hospitals.

Two features of the PfP campaign, the national focus and the robust implementation of its multiple partnership strategies, both key to achieving national patient safety goals, coupled with a variety of data and implementation constraints, limit the ability to capture the impact of the PfP campaign in the quantitative analyses. Chapter 7 in the report provides a discussion of potential reasons why these results occurred. Still,

[8-14] The PDSA cycle was derived from the work of Walter Shewhart and his student W. Edwards Deming. Additional background can be found here at https://www.deming.org/theman/theories/pdsacycle.

despite the uncertainty presented by these disparate results, the lessons learned over the four years of the PfP campaign point toward several areas of opportunity for improvement in the design and implementation of large-scale quality improvement initiatives.

The qualitative results identify several aspects of the PfP campaign that should be considered for continuation in future quality improvement efforts. Characteristics such as peer-to-peer networking; combining federal, state, local, and private partners; and instilling a new culture of collaboration in quality improvement have the potential to combine and interact in ways that amplify the ability of the hospital networks to communicate information quickly and support an active learning community. Within such a community, providers are likely to experience a greater exposure to resources, experience, and support than might otherwise be available to a provider not participating in the network.

In addition to these aspects of the campaign network structure, participants also indicated that the alignment of several mechanisms such as payment incentives, public reporting for transparency, and setting bold aims may work together to motivate transformation in ways that would not be as achievable in the absence of such alignment. To the extent that the factors motivating quality improvements can be brought into alignment with other aspects of health care delivery and performance assessments, the competition among key drivers for change will be reduced and the likelihood of achieving the desired goals increased.

Despite the qualitative evidence of the campaign's success, quantitative evidence supporting the impact of the HEN component of the PfP campaign did not provide clear evidence of an impact above what might have been achieved in the absence of the campaign. The reasons for this lack of evidence regarding campaign impacts are varied, and include issues of spillover in HEN intervention activities, competing quality improvement initiatives associated with both HEN-aligned and non-HEN-aligned hospitals, and the limited ability to link specific intervention activities with improvements in outcomes observed. Future large-scale campaigns to improve patient safety and care can make use of the lessons learned in the PfP campaign, to improve implementation and better achieve intended goals.

Designing quality improvement initiatives intended to operate on a national scale is a challenging task, one that requires the flexibility of permitting treatment to adapt to local needs, balanced against the desire to provide rigorous evaluation analyses. Drawing from literature on complex adaptive systems, each large scale system can be described as a network of organizations that are working together toward achieving a goal. Development of the system would benefit from an approach that considers ways in which the functioning of each specific part of the system might be measured and evaluated. The totality of evaluative activities can then be used to better understand the functioning of the system and provide proper attribution of improvements where needed. The use of these principles to guide the design of future large-scale campaigns will be an enduring legacy of the PfP campaign.

The PfP campaign was implemented at the beginning of a period during which rates of patient harm decreased substantially on a national basis—a fact that has been documented well by the evidence. After a decade of little or no response to the challenge posed by the Institute of Medicine in its 1999 study, *To Err is Human*, the past 4 to 5 years have seen real and substantial improvements in patient safety throughout the United States. Whether a portion of this decline in patient harms can be precisely and directly attributed to the PfP campaign is not clear at this time; the evidence is mixed and inconclusive. However, the participants in the campaign generally support the goals of the campaign and the manner in which activities drove providers toward those goals. The past four years of the PfP campaign represent the first turn through Deming's PDSA cycle. The lessons learned will be invaluable for improving the process in future cycles.

9. References

Agency for Healthcare Quality and Research. Comorbidity Software. Version 3.7. Available at: https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp. Accessed on: March 15, 2015.

Agency for Healthcare Research and Quality. Efforts To Improve Patient Safety Result in 1.3 Million Fewer Patient Harms: Interim Update on 2013 Annual Hospital-Acquired Condition Rate and Estimates of Cost Savings and Deaths Averted From 2010 to 2013 (Publication No. 15-0011-EF). Available at: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2013.html.

Agency for Healthcare Research and Quality (AHRQ). Interim Update on 2013 Annual Hospital-Acquired Condition Rate and Estimates of Cost Savings and Deaths Averted From 2010 to 2013. Available at http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/pfp/interimhacrate2013.pdf . Accessed on: August 29, 2015.

Agency for Healthcare Research and Quality (AHRQ). Making Health Care Safer II: An Updated Critical Analysis of the Evidence for Patient Safety Practices. Evidence Report/Technology Assessment Number 211 (AHRQ Publication No. 13-E001-EF). Accessed in March 2013.

Agency for Healthcare Research and Quality. Partnership for Patients. Available at: http://www.ahrq.gov/professionals/quality-patient-safety/pfp/index.html.

Agency for Healthcare Research and Quality (AHRQ). Advances in Patient Safety: From Research to Implementation. Volume 4: Programs, Tools, and Products (AHRQ Publication No. 05-0021-4).

Bates DW, Spell N, Cullen DJ, et al. The Costs of Adverse Drug Events in Hospitalized Patients. *Journal of the American Medical Association*. 1997; 277:307–311.

Bell CM, Redelmeier DA. Mortality among Patients Admitted to Hospitals on Weekends as Compared with Weekdays. *New England Journal of Medicine*. 2001; 345:663-668.

Burns L, Pauly M. Integrated Delivery Networks: A Detour on the Road to Integrated Health Care? *Health Affairs*. 2002; 21(4):128-143.

Busso M, DiNardo J, McCrary J. (2014). New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators. T*he Review of Economics and Statistics*. December 2014; 96 (5): 885-897.

Cattaneo MD. Efficient Semiparametric Estimation of Multi-Valued Treatment Effects Under Ignorability. *Journal of Econometrics.* 2010; 155(2):138-154.

Chassin MR, Loeb JM. High-Reliability Health Care: Getting There from Here. *Milbank Quarterly. 2013; 91(3):*459-490.

Chen A, Kranker K, Zurovac J, et al. *Project Evaluation Activity in Support of Partnership for Patients: Evaluation Progress Report, I*. Report Submitted to the Centers for Medicare and Medicaid Services. Phoenix: Health Services Advisory Group, and Washington, NJ: Mathematica Policy Research, December 2014.

Clark SL, Frye DR, Meyers JA, et al. Reduction in Elective Delivery at < 39 Weeks of Gestation: Comparative Effectiveness of 3 Approaches to Change and the Impact of Neonatal Intensive Care Admission and Stillbirth. *American Journal of Obstetrics & Gynecology*. 2010.

Clarkwest A, Kranker K, Witmer S, et al. Project Evaluation Activity in Support of Partnership for Patients: Task 2 Evaluation Plan. Submitted to the Centers for Medicare & Medicaid Services. Phoenix: Health Services Advisory Group and Washington, NJ: Mathematica Policy Research, April 2012.

Classen DC, Resar R, Griffin F, et al. "Global Trigger Tool" Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured. *Health Affairs*. 2011; 30(4):581–589.

Damschroder LJ, Aron, DC, Keith RE, et al. Fostering Implementation of Health Services Research Findings into Practice: a Consolidated Framework for Advancing Implementation Science. *Implementation Science.* 2009; 4:50.

Daughtridge GW, Archibald T, Conway P. Quality Improvement of Care Transitions and the Trend of Composite Hospital Care. *Journal of the American Medical Association.* 2014; 311(10):1013-1014.

Dicuccio MH. The Relationship Between Patient Safety Culture and Patient Outcomes: A Systematic Review. *Journal of Patient Safety*. 2015; 11(3):135-142.

DuGoff EH, Schuler M, Stuart EA. Generalizing Observational Study Results: Applying Propensity Score Methods to Complex Surveys. *Health Services Research.* 2014; 49(1):284-303.

ECRI Institute. Top 10 Patient Safety Concerns for Healthcare Organizations 2015. ECRI Institute. Available at: https://www.ecri.org/Pages/Top-10-Patient-Safety-Concerns.aspx. Accessed on: August 29, 2015.

Elixhauser A, Steiner C, Harris DR, et al. Comorbidity Measures for Use with Administrative Data. *Medical Care*. 1998; 36(1):8-27.

Encinosa W, Hellinger F. The Impact of Medical Errors on 90-Day Costs and Outcomes: An Examination of Surgical Patients. *Health Services Research*. 2008; 43:2067–2085.

Farley JF, Harley CR, Devine JW. A Comparison of Comorbidity Measurements to Predict Healthcare Expenditures. *American Journal of Managed Care.* 2006; 12(2):110-119.

Fixsen DL, Naoom SF, Blase KA, et al. *Implementation Research: A Synthesis of the Literature FMHI*. (Publication No. 231). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network. 2005.

Freeman S. Capacity and Utilization in Health Care: The Effect of Empty Beds on Neonatal Intensive Care Admission, Social Science Research Network, March 2015. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2456305. Accessed on: April 10, 2015.

Gelman A, Carlin JB, Stern HS, et al. *Bayesian Data Analysis*. 3rd ed. London: Chapman & Hall/CRC; 2013.

Goldstein SD, Papandria DJ, Aboagye J, et al. The "Weekend Effect" in Pediatric Surgery—Increased Mortality for Children Undergoing Urgent Surgery During the Weekend. *Journal of Pediatric Surgery.* 2014; 49:1087-1091.

Gorodeski EZ, Starling RC, Blackstone EH. Are All Readmissions Bad Readmissions? *New England Journal of Medicine*. 2010; 363(3):297-298.

Greenhalgh T, Robert G, Macfarlane F, et al. Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. *Milbank Quarterly*. 2004; 82(4):581–629.

Guo S, Fraser MW. *Propensity Score Analysis: Statistical Methods and Applications*. Los Angeles, CA: SAGE Publications; 2010.

Heidenrieich PA, Sahay A, Kapoor JR, et al. Divergent Trends in Survival and Readmission Following a Hospitalization for Heart Failure in the Veterans Affairs Health Care System 2002 to 2006. *Journal of the American College of Cardiology*. 2010; 56(5):362-368.

Himmelstein D, Woolhandler S. Quality Improvement: 'Become Good at Cheating and You Never Have to Become Good At Anything Else.' *Health Affairs Blog*, August 27, 2015. Available at: http://healthaffairs.org/blog/2015/08/27/quality-improvement-become-good-at-cheating-and-you-never-need-to-become-good-at-anything-else. Accessed on: August 28, 2015.

Hirano K, Imbens GW. Estimation of Causal Effects using Propensity Score Weighting: An application to Data on Right Heart Catheterization. *Health Services and Outcomes Research Methodology*. 2001; 2(3):259-278.

Health Services Advisory Group and Mathematica Policy Research. Project Evaluation Activity in Support of Partnership for Patients: Evaluation Annual Report. Submitted to the Centers for Medicare and Medicaid Services. Phoenix: Health Services Advisory Group and Washington, DC: Mathematica Policy Research, September 2014.

Imbens GW. The role of the Propensity Score in Estimating Dose-Response Functions. *Biometrika*. 2000; 87(3):706-710.

Kaissi A, Begun J. Fads, Fashions, and Bangwagons in Health Care Strategy. *Health Care Management Review*. 2008; 33(2):94-102.

Kandilov AMG, Coomer NM, Dalton K. The Impact of Hospital-Acquired Conditions on Medicare Program Payments. *Medicare and Medicaid Research Review*. 2014; 4(4).

Klass P. Death Takes a Weekend. *New England Journal of Medicine.* 2015; 372:402-405.

Kreuter MW, McBride TD, Caburnay CA, et al. What Can Health Communication Science Offer for ACA Implementation? Five Evidence-informed Strategies for Expanding Medicaid Enrollment. *Milbank Quarterly*. 2014; 92(1):40–62.

McFadden KL, Stock GN, Gowen CR. Leadership, Safety Climate, and Continuous Quality Improvement: Impact on Process Quality and Patient Safety. *Health Care Management Review*. 2015; 40(1):24-34.

Nápoles AM, Santoyo-Olsson J, Stewart AL. Methods for Translating Evidence-Based Behavioral Interventions for Health-Disparity Communities. *Preventing Chronic Disease*. 2013; 10:E193.

National Quality Forum (NQF). Safe Practices for Better Healthcare—2009 Update, A Consensus Report. Washington, DC: National Quality Forum, March 2009. Available at: https://www.qualityforum.org/Publications/2009/03/Safe_Practices_for_Better_Healthcare%E2%80%932009_Update.aspx. Accessed on: August 28, 2015.

Nichols A (2008). Erratum and Discussion of Propensity–Score Reweighting. *Unpublished manuscript*.

Office of Inspector General, U.S. Department of Health and Human Services. Adverse Events In Hospitals: National Incidence Among Medicare Beneficiaries (Publication 2010b). Available at: http://oig.hhs.gov/oei/reports/oei-06-09-00090.pdf. Accessed on: August 13, 2015.

Peberdy MA, Ornato JP, Larkin GL, et al. Survival from In-Hospital Cardiac Arrest During Nights and Weekends. *Journal of the American Medical Association*. 2008; 299:785-792.

Pentland D, Forsyth K, Maciver D, et al. Key Characteristics of Knowledge Transfer and Exchange in Healthcare: Integrative Literature Review. *Journal of Advanced Nursing.* 2011; 67(7):1408–1425.

Perla RJ, Bradbury E, Gunther-Murphy C. Large-Scale Improvement Initiatives in Healthcare: A Scan of the Literature. *Journal of Healthcare Quality.* 2013; 35(1):30–40.

Powell BJ, McMillen JC, Proctor EK, et al. A Compilation of Strategies for Implementing Clinical Innovations in Health and Mental Health. *Medical Care Research and Review*. 2012; 69(2):123–157.

Qualidigm. Medicare Patient Safety Monitoring System (MPSMS). 2012 Annual Report of 2011 Data, Overview. Available at: http://www.qualidigm.org/wp-content/uploads/2012/02/MPSMS-Overview-Final-Report.pdf. Accessed on: July 21, 2015.

Reason, J. *Human Error*. Cambridge, MA: Cambridge University Press; 1990.

Rivard PE, Luther SL, Christiansen CL, et al. Using Patient Safety Indicators to Estimate the Impact of Potential Adverse Events on Outcomes. *Medical Care Research and Review*. 2008; 65:67-87.

Rouse, WB. Health Care as a Complex Adaptive System: Implications for Design and Management. *The Bridge*. 2008; 38(1):17-25.

RTI International. Health IT Safety Center Roadmap. Prepared for the Office of the National Coordinator for Health Information Technology. NC: RTI, July 2015. Available at: http://www.healthitsafety.org/. Accessed on: August 28, 2015.

Sammer CE, Lykens K, Singh K, et al. What is Patient Safety Culture: A Review of the Literature. *Journal of Nursing Scholarship.* 2010; 42(2):156–165.

Schneeweiss, Sebastian, Maclure M. Use of Comorbidity Scores for Control of Confounding in Studies Using Administrative Databases. *International Journal of Epidemiology.* 2000; 29(5):891-898.

Shafer T, Wagner D, Chessare J, et al. US Organ Donation Breakthrough Collaborative Increases Organ Donation. *Critical Care Nursing Quarterly*. 2008; 31(3):190-210.

The American Hospital Association/Health Research & Educational Trust Hospital Engagement Network. Annual Report – December 2013. Available at: http://www.aha.org/content/14/ahahrethen-report-final.pdf. Accessed inJuly 2015.

Tsai TC, Joynt KE, Oray EJ, et al. Variation in Surgical-Readmission Rates and Quality of Hospital Care. *New England Journal of Medicine*. 2013; 369(12): 1134-1142.

van Walraven C, Austin PD, Jennings A., et al. A Modification of the Elixhauser Comorbidity Measures Into a Point System for Hospital Death Using Administrative Data. *Medical Care*. 2009; 47(6):626-633.

Wang Y, Eldridge N, Metersky ML, et al. National Trends in Patient Safety for Four Common Conditions, 2005–2011. *New England Journal of Medicine. 2014; 370:*341-351.

Weick K, Sutcliffe K. Managing the Unexpected: Resilient Performance in an age of Uncertainty. San Francisco, CA: Jossey Bass.

What Works Clearinghouse. Procedures and Standards Handbook. Version 2.1. Available at: http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf. Accessed on: September 16, 2015.

Yuan CT, Nembhard IM, Stern MF, et al. "Blueprint for the Dissemination of Evidence-Based Practices in Health Care." Commonwealth Fund, Pub. 1399, Vol. 86, May 2010

Zhan C, Friedman B, Mosso A, et al. Medicare Payment for Selected Adverse Events Under the Prospective Payment System: Building the Business Case for Investing in Patient Safety Improvement. *Health Affairs*. 2006; 25:1386-1393.

Zhan C, Miller M. Excess Length of Stay, Charges, and Mortality Attributable to Medical Injuries During Hospitalization. *Journal of the American Medical Association*. 2003; 290:1868–1874.

Zingg W, Holmes A, Dettenkofer M, et al. Hospital Organization, Management, and Structure for Prevention of Health-Care-Associated Infection: A Systematic Review and Expert Consensus. *The Lancet.* 2015; 15(2):212-224.