Center for Medicare and Medicaid Innovation
Contract No. GS-10F-0166R
Task Order: HHSM-500-2011-00153G

Project Evaluation Activity in Support of Partnership for Patients: Task 2 Evaluation Progress Report—Appendices

February 2014
Submitted 7/10/2014

HSAG HEALTH SERVICES ADVISORY GROUP
MATHEMATICA Policy Research

CONTENTS

A. Appendix A: Methods A-1
B. Appendix B: Data Sources B-1
C. Appendix C: Summary of Measure Limitations C-1
D. Appendix D: AHA/HRET Initial Analyses of Engagement and Outcomes D-1

Are Hospitals More Engaged in Quality Improvement Efforts More Successful? An Assessment of HRET HEN Hospitals over 26 Months Executive Summary—April 28, 2014 ..... D-1

Are Hospitals More Engaged in Quality Improvement Efforts More Successful? An Assessment of HRET HEN Hospitals over 26 Months—April 25, 2014 ..... D-5

APPENDIX A: METHODS

The analyses in the report examine the following:

- Change in Rate: How has the incidence of the targeted outcomes changed during the course of the Parternship for Patients (PfP) campaign? That is, have adverse events and readmissions become less common?
- Change in Speed with which Rate is Changing: If rates of targeted events were changing before PfP activities began, did the trend shift after activities began?
- Differences in improvement between Hospital Engagement Networks (HENs)-aligned and non-aligned comparison hospitals: Have outcomes improved more among hospitals that worked with a HEN than among similar hospitals that did not work with a HEN?

The results in the report are based on the following analytic approaches:

- Change in Rate:
  - Statistical process control charts
  - *t*-tests of difference in incidence between baseline and follow-up
- Change in Speed with which Rate is Changing: t-tests of difference in trajectory of change between baseline and follow-up.
- Differences between HEN-aligned and non-aligned comparison hospitals: Difference-in-difference comparisons of change in the incidence of targeted outcomes of HEN-aligned and a comparison group of non-aligned hospitals, with propensity score reweighting used to create the comparison group.

The sections below provide detail on each of the analytic approaches used.

Statistical Process Control Charts

The Evaluation Contractor used two types of Shewhart control charts to examine changes in the incidence of targeted outcomes.[A-1]

- p′ charts: for outcomes measured as rates
- *X*-charts: for National Healthcare Safety Network (NHSN) standardized infection ratio (SIR) measures

The control limits for all charts are created based on the earliest year of data available. The p′ chart control limits are calculated using a sigma (σ_p) on the standard error of the theoretical distribution of the probability of the event ($\bar{p}$) for that year period, with an adjustment for over-dispersion that commonly occurs with large administrative data sets. That is, sigma is calculated as:

[A-1] For more details, see Ryan, Thomas. P. Statistical Methods for Quality Improvement. New York: John Wiley & Sons, 1989, or Provost, Lloyd P., and Sandra Murray. The Health Care Data Guide: Learning from Data for Improvement. San Francisco: Jossey-Bass, 2011.

$$\sigma_{\bar{p}} = \sqrt{\frac{\bar{p} * (1 - \bar{p})}{n}}$$

where n is the average number of denominator cases for each measurement period. The control limits are calculated as 3 sigmas to either side of the rate over the year period. Those theoretically-based limits are then adjusted for over-dispersion based on the observed average moving range ($\overline{MR}$) over the first year's data points (see Provost and Murray 2011).[A-2]

The control limits of the *X*-charts are calculated as $\bar{x} \pm 2.66\overline{MR}$, with both $\bar{x}$ and $\overline{MR}$ being calculated using the first year of observed data points.

Rules for Detecting Special Cause Variation in Shewhart Charts

The following rules were used to detect special cause variation in the control charts:

1. One or more points outside the control limits
2. Two or more consecutive points between 2 and 3 sigma limits from the center line
3. A run of 8 or more consecutive points above or below the center line
4. Six or more consecutive points increasing or decreasing

A chart meeting any of those criteria is deemed to reflect evidence of non-random variation in the outcome. Specifically, the chart constitutes evidence of a systematic change occurring in the incidence of the outcome.[A-3]

***t*-tests of Change in Rate and Change in Speed with which Rate is Changing**

The team used *t*-tests to examine both change in rate and change in the speed with which the rate is changing. To examine change in rates, the tests compared the rate (or other incidence) measure in each follow-up quarter to the rate at baseline. The outcomes are considered statistically different if the *p*-value from the two-sided *t*-test is less than 0.05. In most cases, the baseline used for tests of differences in rates is the first year of data available. In a couple of cases a shorter baseline is used when data are available for only a short time horizon. Those include the t-tests using data reported by HENs and the NHSN-based catheter-associated urinary tract infections (CAUTI) and surgical site infection (SSI) measures. For HEN-reported early elective delivery (EED) and readmissions measures which use cohorts with consistent baselines, the Evaluation Contractor used February 2012 to July 2012 data as the baseline. For other HEN-reported measures presented in the report

---

A-2 This is done by standardizing the individual probabilities for each period, i, then calculating a sigma for the standardized probabilities based on the $\overline{MR}$ of those probabilities as $\sigma_z = \overline{MR}/1.128$. The center line in the adjusted chart remains the same as in the unadjusted chart ($\bar{p}$), but the control limits are modified as equal to $\pm 3\sigma_{\bar{p}}\sigma_z$ around that center line, rather than $\pm 3\sigma_{\bar{p}}$.

A-3 The analyses presented here do not examine change in the variance of the outcome, though that can be examined using Shewhart charts as well.

the baseline period varies across HENs depending on the time horizon over which the particular HEN reported data. The time periods are predominantly 2010 or 2011, and are always at least 3 months in length. Given that the NHSN CAUTI and SSI measures (standardized infection ratios [SIRs]) are reported only starting in Q1 2012, that quarter is used as the baseline, rather than using the whole of 2012 as baseline.

To test for change in the speed with which the rate is changing, the Evaluation Contractor performed t-tests of the differences in the total change in incidence measures over the earliest annual period of data available, compared to annual change occurring during the follow-up period. For example, the National Database of Nursing Quality Indicators (NDNQI®) rates are reported from Q1 2011 through Q3 2013.[A-4] So the baseline measure is the change in rates between Q1 2011 and Q1 2012. The t-tests compare that change to the change over an annual follow-up period such as Q3 2012 to Q3 2013 (the latest year of follow-up data) or Q12012 to Q1 2013 (the earliest year of follow-up data).

Comparison Group Analyses

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-aligned hospitals and compared change in the outcomes between the two groups. The comparison group was created using propensity score reweighting. Regression-based difference-in-differences analyses were used to compare changes between the groups. The comparison group analyses could be performed only for the measures based on Medicare claims or NHSN data.[A-5] For other data sources data on non-aligned hospitals were not available or it was not possible to construct a rigorous comparison group using the data for the non-aligned hospitals. The sections below describe the propensity score reweighting and difference-in-differences approaches used by the Evaluation Contractor.

There are two caveats to the comparison group analysis. First, as discusses in the report, the HENs' work is only one of the elements used to achieve PfP goals; thus the comparison group analyses only address the effectiveness of the HEN activities. Second, to the extent that non-aligned as well

---

A-4 National Database of Nursing Quality Indicators (NDNQI®) is a registered trademark of the ANA.

A-5 Analyses with NHSN data were conducted with data provided to the Evaluation Contractor by staff at the Centers for Disease Control (CDC). The Evaluation Contractor did not have access to patient- or hospital-level NHSN data. Instead, the Evaluation Contractor provided detailed specifications to the CDC, who aggregated the SIRs based on the Team's specifications and provided aggregated quarterly SIRs and confidence intervals. The CDC provided aggregate SIRs for (1) discharges occurring in HEN-aligned and a comparison group of non-aligned hospitals, respectively (CAUTI and central line-associated blood stream infection [CLABSI] only); and (2) discharges in smaller subgroups of hospitals. For example, the CDC reported quarterly standardized infection ratios (SIRs) for discharges occurring in HEN-aligned rural hospitals. The Evaluation Contractor estimated impacts on patient outcomes by using the aggregated SIRs reported by the CDC, as described below. The Evaluation Contractor also combined SIRs for the unweighted set of non-aligned hospitals, which the CDC provided to the Evaluation Contractor, with the SIRs for HEN-aligned hospitals to create quarterly national measures for the trend analyses.Although the impact analysis techniques used in this report share the same fundamental approach, comparing the changes observed in the adverse event and readmission rates in HEN-aligned hospitals with changes observed in a comparison group of similar hospitals, there are differences in the details depending on whether the data are discharge-level or aggregated across hospitals. The main advantage of discharge-level regression models that can be used with Medicare data is that the data allow for potential correlations across patients. Both the discharge-level and hospital-level analyses permit regression-adjustment for factors that cannot be included in the analyses using the aggregated NHSN SIRs.

as HEN-aligned hospitals received benefits from PfP, the comparison method may underestimate the true impact of PfP.

Propensity Score Reweighting

For purposes of understanding the effect of HENs' efforts, it is important to assemble a comparison group that is similar to the group of HEN-aligned hospitals in order to understand how different HEN-aligned hospitals would have been had the hospitals not worked with a HEN. A simple comparison of observed improvement in HEN-aligned versus non-aligned hospitals will not serve that purpose because hospitals that elect to work with a HEN differ in important ways from those that do not, and differences in outcomes might result from underlying differences in hospital mix rather than from the effects of PfP.

Using a statistical technique called propensity score reweighting, the Evaulation Contractor created a comparison group from the pool of non-aligned hospitals. Propensity score reweighting produces a comparison group of non-aligned hospitals that is similar to HEN-aligned hospitals on observable characteristics of hospitals and their patients, by assigning different weights to non-aligned hospitals depending on their similarity to HEN-aligned hospitals, giving more weight to non-aligned hospitals that are more similar to HEN-aligned hospitals and less weight to non-aligned hospitals that are less similar.

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

The Evaluation Contractor estimated a logistic (logit) regression model using the predictor variables in Table A-1. The baseline hospital characteristics used to create the weights were drawn from data from the 2010 American Hospital Association (AHA) Annual Survey and Medicare claims from the pre-campaign years (2009 and 2010). Examples of characteristics from the AHA survey include size, urbanicity, and whether the hospital has an electronic health record (EHR) system. The claims data provided data on level and trend in the adverse event and readmission rates before the start of PfP, as well as the demographic composition of the patients served.

The predictor variables included pre-intervention rates and trends in the outcome measures. Other variables, such as region, urbanicity, and hospital size are characteristics included as broad differentiators of hospitals and their contexts. A further set of items was included based on having been found to predict HEN alignment in the Evaluation Contractor's baseline analysis. Examples include teaching hospitals or the percentage of physicians who are intensivists. The predictor variables were entered into the logit model as predictors of treatment status, defined as a binary dependent variable that equals one for HEN-aligned (treatment) hospitals and zero for non-aligned hospitals. Hospitals were considered to be HEN-aligned if they appeared on a HEN's roster of hospitals in June 2012. Separate propensity models were estimated for each adverse event area (AEA) and for readmissions. To maintain continuity with the impact regressions (in which discharges are the unit of analysis and thus hospitals implicitly are weighted by the number of

relevant discharges), hospitals were weighted by the number of discharges in the measure denominator in 2010.

For the analyses with NHSN data, the Evaluation Contractor provided propensity score weights to the Centers for Disease Control and Prevention (CDC) staff, who used the weights when calculating the SIRs. This reweighting approach was intended to make the comparison group as similar as possible to the HEN-aligned group, although the process was less effective for some NHSN outcomes—CAUTI in particular—because the Evaluation Contractor did not have access to baseline hospital-level data to use in the propensity score model and had to rely on Medicare data alone when calculating the weights.

Table A-1—Variables Included in the Propensity Score Models, by AEA

Variable	Central Line-Associated Blood Stream Infection	Falls	Pressure Ulcers	Venous Thromboembolism	Catheter-Associated Urinary Tract Infection	Surgical Site Infection	Readmissions
Hospital Characteristics[a]							
Bed size[b]	✓	✓	✓	✓	✓	✓	✓
Ownership type[c]	✓	✓	✓	✓	✓	✓	✓
Has EHR system[d]	✓	✓	✓	✓	✓	✓	✓
Census region[i]	✓	✓	✓	✓	✓	✓	✓
Urban/rural type[e]	✓	✓	✓	✓	✓	✓	✓
Teaching hospital	✓	✓	✓	✓	✓	✓	✓
Hospital belongs to a network	✓	✓	✓	✓	✓	✓	✓
Hospital belongs to health care system	✓	✓	✓	✓	✓	✓	✓
Rural referral center	✓	✓	✓	✓	✓	✓	✓

Table A-1—Variables Included in the Propensity Score Models, by AEA							
Variable	**Central Line-Associated Blood Stream Infection**	**Falls**	**Pressure Ulcers**	**Venous Thromboembolism**	**Catheter-Associated Urinary Tract Infection**	**Surgical Site Infection**	**Readmissions**
Intensivist, as percentage of total physicians	✓	✓	✓	✓	✓	✓	✓
Inpatient prospective payment system (IPPS) hospital							✓
Critical Access Hospital (CAH)							✓
Community based care transition program (CCTP) participation							✓
Patient Case Mix Characteristics (Hospital Level)[a]							
Mean patient age at baseline	✓	✓	✓	✓	✓	✓	✓
Race/ethnic composition of inpatient population[f]	✓	✓	✓	✓	✓	✓	✓
Percentage of patients who are female	✓	✓	✓	✓	✓	✓	✓
Pre-Intervention Outcomes[a]							
2010 adverse event rate[g]	✓	✓	✓	✓	✓	✓	

Table A-1—Variables Included in the Propensity Score Models, by AEA							
Variable	**Central Line-Associated Blood Stream Infection**	**Falls**	**Pressure Ulcers**	**Venous Thromboembolism**	**Catheter-Associated Urinary Tract Infection**	**Surgical Site Infection**	**Readmissions**
Percentage point difference in adverse event rate between 2009 and 2010[g]	✓	✓	✓	✓	✓	✓	
2010 30-day readmission rates for all causes, acute myocardial infarction (AMI), heart failure (HF), and pneumonia (PN) (four measures)[h]							✓
Percentage point difference in readmission rates between 2009 and 2010 for all causes, AMI, HF, and PN[h]							✓

Notes: Propensity score models for CLABSI, CAUTI, and SSI were also created for hospitals reporting to the NHSN. Patient characteristics and historical outcomes data from Medicare claims were used to estimate the scores because the data on patient characteristics and historical outcomes were not available from NHSN. The NHSN propensity score models for each AEA (CLABSI, CAUTI, or SSI) used the same set of covariates as in the Medicare Patient Safety Indicator (PSI)/Hospital-Acquired Condition (HAC) models for the corresponding AEA. The sample of hospitals; however, differed somewhat, consisting of only the subset of those hospitals that also reported to NHSN throughout the reporting period.

[a]The 2010 AHA annual survey was the data source for the hospital characteristics. Patient characteristics and pre-intervention outcome variables were derived from Medicare claims.

[b]Bed size was categorized as fewer than 100 beds (non-critical access hospitals [CAH]); 100–199 beds (non-CAH), 200–399 beds (non-CAH); or more than 400 beds (non-CAH). CAHs were entered as a separate category in the analyses of readmission rates. No such variable is included in the adverse events because CAHs were excluded from those analyses given the lack of consistent reporting of present-on-admission indicators by CAHs and other hospitals that are not part of the IPPS.

[c]Ownership types included investor-owned (for-profit), non-government not-for-profit, federal government, and non-federal government. Given the smaller hospital samples included in the analyses, the variable for federal government ownership was not included in the falls and PrU propensity score models.

[d]Full, partial, or no adoption of EHRs. This variable was missing for a substantial portion of the sample. A "missing data" indicator was used in the model to avoid dropping the entire case from the regression model.

[e]Whether the hospital was located in a rural county was determined by the Rural-Urban Commuting Area (RUCA) code. For hospitals not in rural areas, the type (size) of urban area was determined by the Core Based Statistical Area (CBSA) code (found in the AHA survey). With that approach, hospital location was identified as rural, metropolitan, micropolitan, or division.

[f]Race/ethnicity was categorized into four mutually exclusive, collectively exhaustive categories: Hispanic, black non-Hispanic, white non-Hispanic, and other non-Hispanic.

[g]For each AEA, hospitals were categorized into one of three groups: zero, low, and high. Zeros had a rate of zero, lows had a rate within the lower half of the distribution of non-zero observations, and highs had a rate in the upper half. A dummy variable for each

of these categories (zeros, low and high) was entered into the model and interacted with the 2010 adverse event rate and percentage point difference in the AEA rate between 2009 and 2010; including the dummy variables and interactions in the propensity score model resulted in an additional nine variables in total.

[h]Four readmission measures were entered into the propensity score model for readmissions: the 30-day all-cause binary readmission rate (main measure as specified in Section II.A.), and the 30-day readmission rates following discharges from stays for AMI, HF, and PN. For each readmission outcome, data were split into three groups: low, medium, and high. Lows had a rate in the lower third of the distribution, medium had a rate within the middle third, and highs had a rate in the upper third. A dummy variable for each of these categories (low, medium, high) was entered into the model and interacted with the 2010 readmission rate and percentage point difference in the rate between 2009 and 2010. Nine variables were included into the propensity score model for the dummy variables and interactions for each of the four readmissions variables, resulting in a total of 36 variables (minus a few variables that dropped out due to collinearity).

[i]Census regions and their component states are: Pacific (CA, OR, and WA), Mountain (AZ, CO, ID, MT, NM, NV, UT, and WY), West South Central (AR, LA, OK, and TX), East South Central (AL, KY, MS, and TN), South Atlantic (DE, DC, FL, GA, MD, NC, SC, VA, and WV), West North Central (IA, KS, MN, MI, NE, ND, and SD), East North Central (IN, IL, MI, OH, and WI), Mid Atlantic (NJ, NY, and PA), and New England (CT, MA, ME, NH, RI, and VT). Hospitals located in United States territories that are located outside the 50 states and DC are categorized as "Associated Areas."

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 Centers for Medicare & Medicaid Services [CMS] quality improvement efforts underway at the same time as PfP).[A-6]

As discussed above, the difference-in-differences **analyses using adverse event measures from Medicare claims** were conducted with patient discharges as the unit of analysis. For each outcome, the sample was limited to hospital discharges that were applicable ("at risk") for the given adverse event—that is, the "denominator" for estimating a particular adverse event rate. Discharges in HEN-aligned hospitals received a weight of one, and discharges in comparison group hospitals received the propensity score–based weight assigned to the hospital where the discharge occurred.

[A-6] Time-invariant characteristics include factors such as region or ownership type. To the extent that patient populations served by hospitals tend to remain relatively stable over time, those differences in those populations will also be adjusted for.

The difference-in-differences regression specification has the following form:

$$y_i = \delta\,(PFP_h * Post_t) + \gamma\, t_t + \phi\, w_i + \theta\, x_i + \beta\, z_h + \varepsilon_{i,}$$

where the outcome variable, y_i, is measured for a hospital discharge (i) occurring in quarter *t* in hospital *h*. The variable PFP_h is a dummy variable for whether or not the hospital where the discharge occurred was aligned with a HEN as of June 2012; $Post_t$ is a dummy for whether or not the discharge occurred after PfP start date; 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 patient demographics (w_i) are age, gender, race/ethnicity. The patient risk factors (x_i) are comorbidities specific for each outcome variable and were chosen in accordance with the risk factors used by the PSI algorithm for calculating risk-adjusted adverse event rates.[A-7] Table A-2 includes a list of the variables used in the Medicare difference-in-differences analyses, by outcome.

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 with the usual properties. Equation (1) was estimated with linear probability models. Compared to nonlinear models, the linear probability model offers three advantages: permitting the use of hospital fixed effects, allowing the marginal effect of the interaction terms to be interpreted without making distributional assumptions, and reducing the computational run times.[A-8]

The coefficient on the first interaction term (PFP_h * $Post_t$) is the impact estimate.[A-9] It captures how the change in an outcome among hospitals that signed up to work with a HEN differs from the change in that outcome among non-aligned hospitals—holding constant differences between hospitals' outcomes at baseline, differences in the characteristics of patients served, differences in stable hospital characteristics that could influence change in outcomes, and external factors that could influence changes over time in outcomes across hospitals in both groups. Subgroup analyses were conducted by introducing interaction terms into equation (1), where the (PFP_h * $Post_t$) term was interacted with a variable denoting the subgroup. Subgroup analyses with hospital characteristics (such as the urban-rural subgroup analyses) also included a term for each subgroup interacted with $Post_t$ to account for differential secular trends in the non-aligned hospital subgroups.

The regression-adjustment for the time-trend graphical analysis, discussed above, is based on regression models similar to equation (1). Specifically, the trend analysis was conducted by

A-7 Controls do not include variables that are potentially endogenous. For example, the Evaluation Contractor did not include large arrays of dummy variables for diagnosis or procedure codes because an adverse event may cause the need for a "cascade" of diagnoses or follow-up procedures to reduce harm or sustain life. Thus, the Evaluation Contractor used patient-specific control variables that (1) are present on admission, (2) represent procedures originally planned or the cause of the admission, and/or (3) otherwise are not added to the claim in the case of an adverse event.

A-8 The Evaluation Contractor used Stata's robust standard errors (SEs) in all models to account for repeated measures within hospitals and heteroskedasticity.

A-9 There is no uninteracted HEN-alignment indicator, PFP_h, because it would be collinear with the hospital fixed effects.

modifying the main difference-in-differences model (equation 1) to include interaction terms with the quarterly time dummies as follows:

$$y_i = \sum_{t=0}^{T} \delta_\tau \left(PFP_h * 1(t = \tau)\right) + \gamma\, t_t + \phi\, w_i + \varepsilon_{i,}$$

where the variables are defined as in equation (1), and "$1(t = \tau)$" is an indicator function that is used to allow the incremental effect of PfP interventions to vary over each quarter. The array of time-specific dummies (t_t) pick up the change in outcomes in the comparison group for each time period. The coefficients for the interaction of the time dummy and the HEN alignment indicator (δ_t) is the regression-adjusted difference between the HEN-aligned and non-aligned hospitals for each quarter. The predicted values from equation (2) are presented graphically, holding the distribution of the covariates fixed.[A-10] For equation (2), a simpler specification was used without hospital fixed effects and with only the array of basic patient demographics (age, gender, and race/ethnicity). Hospital fixed effects were not included in the model because fixed-effect parameters would have been perfectly correlated with the HEN-alignment dummy (PFP_h), making it statistically infeasible to estimate the counterfactual for the non-aligned group. The graphs allowed the Evaluation Contractor to determine when the largest changes in adverse event rates might have occurred; they also permitted the Evaluation Contractor to explore, for example, whether changes in particular outcomes followed the implementation of particular HEN activities. Differences in the control variables aside, the difference-in-differences estimate from equation (1) may be interpreted as a (weighted) average effect across all quarters in the intervention period (that is, an average of δt over quarters in the post-intervention period) compared to the average difference in the baseline period.

As discussed above, the difference-in-differences **analyses with Medicare readmission measures** used hospital-level data constructed from claims but were otherwise similar to the analyses with adverse events. The sample consisted of one observation for each hospital for each year. Observations for HEN-aligned hospitals received a weight of one, and comparison group hospitals received the propensity score–based weight. The difference-in-differences regression specification had the following form for the readmission outcomes:

$$y_{ht} = \delta\,(PFP_h * Post_t) + \gamma\, t_t + \phi\, \bar{w}_{ht} + \theta\, \bar{x}_{ht} + \beta\, z_h + \varepsilon_{ht,}$$

where the outcome variable, y_{ht}, is measured for a hospital (h) in each year (t). For example, the outcome might be the 30-day all-cause readmission rate, a variable that ranges from 0 to 1 for each hospital in each year. The variables PFP_h, $Post_t$, and t_t were defined as above. The Evaluation Contractor controlled for patient demographics and comorbidities by aggregating them to the hospital-level each year ($\bar{w}_{ht}$ and $\bar{x}_{ht}$, respectively). The Evaluation Contractor continued to include a vector of hospital dummies (z_h). ε_{ht} is an error term with the usual properties.[A-11] As with

[A-10] The distribution of patient characteristics was held constant, using the distribution observed for patients admitted to HEN-aligned hospitals in 2012.

[A-11] Equation (2) was estimated with linear regression models. For all regression models, the Evaluation Contractor adjusted the SEs to account for repeated measures within hospitals and heteroskedasticity (report robust SEs clustered by hospital).

the equations discussed above, the coefficient on the first interaction term (PFP_h * $Post_t$) is the impact estimate. The time-trend and subgroup analyses were also similar to the discharge-level analyses for adverse events but underwent modification to account for the use of hospital-level data instead of discharge-level data.

Difference-in-differences analyses with the NHSN SIR outcomes were calculated in two steps. First, the CDC calculated aggregate SIRs (and 95 percent confidence intervals) for the treatment and comparison groups over time for HEN-aligned and non-aligned hospitals, with hospitals weighted with propensity score–based weights. Second, the Evaluation Contractor estimated the difference-in-differences by using the aggregate results from the CDC. The difference-in-difference estimates for the effect of PfP on the standardized infection rates from CLABSI and CAUTI are as follows:

$$\Delta_T = T_1 - T_0$$
$$\Delta_C = C_1 - C_0$$
$$DID = \Delta_T - \Delta_C$$

where T_0 is the average SIR for HEN-aligned hospitals in the baseline period, T_1 is the average SIR for HEN-aligned hospitals in the follow-up period, and C_0 and C_1 are the corresponding SIRs for comparison hospitals in the baseline and follow-up periods. These averages are calculated using the quarterly SIR numerators and denominators across the time periods of interest. For CLABSI, the baseline period is all four quarters of 2011, and the follow-up period is all four quarters of 2012. For the remaining SIR measures, the baseline period is quarter 1 of 2012, and the follow-up period is defined as quarters 2 to 4 of 2012.[A-12]

The standard errors for the difference-in-difference estimates are calculated as follows:

$$\sigma_{\Delta_T} = \sqrt{\sigma_{T_1}^2 + \sigma_{T_0}^2}$$
$$\sigma_{\Delta_C} = \sqrt{\sigma_{C_1}^2 + \sigma_{C_0}^2}$$
$$\sigma_{DID} = \sqrt{\sigma_{\Delta_T}^2 + \sigma_{\Delta_C}^2} = \sqrt{\sigma_{T_1}^2 + \sigma_{T_0}^2 + \sigma_{C_1}^2 + \sigma_{C_0}^2}$$

where σ_{T0} is the standard error of the average SIR for HEN-aligned hospitals in the baseline period, σ_{T1} is the standard error of the average SIR for HEN-aligned hospitals in the follow-up period, and σ_{C0} and σ_{C1} are the corresponding standard errors for comparison hospitals in the baseline and follow-up periods. The standard errors for the baseline and follow-up periods are computed as an equally weighted average of the corresponding quarterly standard errors. This calculation assumes zero covariance between quarters, which is a necessary assumption given the data that were provided. A similar assumption is made in the calculation of the difference-in-

[A-12] Alternative specifications were to define the baseline period as the last quarter before PfP was implemented and to define the post-period using a specific quarter after PfP was implemented. The estimates were generated using all combinations of these definitions and the main definitions of the baseline and follow-up periods.

differences standard errors; the calculation assumes zero covariance between the baseline and follow-up periods, as well as between the HEN-aligned and comparison hospitals. For purposes of understanding the effect of HENs' efforts, it is important to assemble a comparison.

Table A-2—Variables Included in the Medicare Difference-in-Differences Models, by AEA

Variable	Central Line-Associated Blood Stream Infection	Falls	Pressure Ulcers	Venous Thromboembolism	Readmissions
Control Variables[a]					
Hospital fixed effects	✓	✓	✓	✓	✓
Quarterly indicators (dummies)	✓	✓	✓	✓	✓
Patient Characteristics[a]					
Patients' age	✓	✓	✓	✓	✓
Patients' race/ethnicity[b]	✓	✓	✓	✓	✓
Patient's sex	✓	✓	✓	✓	✓
AEA-Specific Risk Factors[c]					
Acquired immune deficiency syndrome (AIDS)			✓	✓	✓
Alcohol abuse				✓	✓
Chronic blood loss anemia	✓			✓	✓

Table A-2—Variables Included in the Medicare Difference-in-Differences Models, by AEA					
Variable	**Central Line-Associated Blood Stream Infection**	**Falls**	**Pressure Ulcers**	**Venous Thromboembolism**	**Readmissions**
Chronic pulmonary disease	✓		✓	✓	✓
Congestive heart failure	✓			✓	✓
Deficiency anemias	✓		✓	✓	
Depression			✓	✓	✓
Diabetes with chronic complications			✓		✓
Diabetes without chronic complications			✓	✓	
Drug abuse	✓		✓	✓	✓
Hypertension	✓		✓	✓	

Table A-2—Variables Included in the Medicare Difference-in-Differences Models, by AEA					
Variable	**Central Line-Associated Blood Stream Infection**	**Falls**	**Pressure Ulcers**	**Venous Thromboembolism**	**Readmissions**
Hypothyroidism			✓	✓	
Liver disease	✓				✓
Lymphoma				✓	✓
Metastatic cancer				✓	✓
No point of origin[d]	✓		✓		✓
No procedure day[e]	✓		✓		✓
Obesity	✓			✓	
Other neurological disorders	✓		✓	✓	

Table A-2—Variables Included in the Medicare Difference-in-Differences Models, by AEA					
Variable	**Central Line-Associated Blood Stream Infection**	**Falls**	**Pressure Ulcers**	**Venous Thromboembolism**	**Readmissions**
Paralysis	✓			✓	✓
Peripheral vascular disease				✓	
Peptic Ulcer with Bleeding					✓
Peripheral vascular disease					✓
Psychoses			✓	✓	✓
Pulmonary circulation disease				✓	✓
Renal failure	✓		✓	✓	✓
Rheumatoid arthritis/collagen vascular disease	✓				
Solid tumor without metastasis			✓	✓	✓

Table A-2—Variables Included in the Medicare Difference-in-Differences Models, by AEA					
Variable	**Central Line-Associated Blood Stream Infection**	**Falls**	**Pressure Ulcers**	**Venous Thromboembolism**	**Readmissions**
Transfer from another acute care facility	✓			✓	✓
Valvular disease	✓			✓	✓
Weight loss	✓		✓	✓	✓

[a]The 2010 AHA annual survey was the data source for the control variables. The Medicare claims file was the data source for the patient characteristics variables.
[b]Race/ethnicity was categorized into four mutually exclusive, collectively exhaustive categories: Hispanic, black non-Hispanic, white non-Hispanic, and other non-Hispanic.
[c]The risk factor indicators were constructed from procedure and diagnosis codes within Medicare claims data by using the PSI software. The Evaluation Contractor used the same risk factors for the PSI measures (CLABSI, falls, pressure ulcers [PrUs], and venous thromboembolism [VTE]) as were used in the PSI software. Risk factors deemed widely relevant on the basis of being included in the CLABSI, PrU, and VTE models were used in the CAUTI and SSI difference-in-differences models.
[d]No point of origin indicates that data were missing in the claims that would otherwise assess where the patient was residing before hospital intake.
[e]No procedure day indicates that the date of the procedure was missing in the claims data.

Estimation of Averted Costs

This section provides more detail about the two available estimates of cost reductions associated with reduced rates of adverse events.

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 and 2012. AHRQ's method is to subtract the national estimate of the number of adverse events in each follow up year (2011 and 2012) from the number estimated for 2010. The difference for each year is multiplied by the cost per event listed in column 1 of Table A-5, developed prior to the start of Partnership for Patients by an interagency HHS group. Please see Table 1-1 in the report body for more about the AHRQ National Scorecard Data used in the estimates. The cost per event estimates were developed prior to PfP by an HHS team including representatives from CMS, AHRQ, CDC, and other agencies; the sources for these estimates are listed in Table A-3. The results for 2011 are provided in Table A-4, and the results for 2012 are provided in Table A-5.

Table A-3—Sources for Cost per Event Used in AHRQ Estimate of Cost Savings

Condition	Cost per Adverse Event	Source
Central Line-Associated Bloodstream Infection (CLABSI)	$17,000	CDC Vital Signs- Central Line Associated Blood Stream Infections- US 2001, 2008, 2009. March 3, 2011 MMWR (e-release March 1, 2011). http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6008a4.htm?s_cid=mm6008a4_w
Venous Thromboembolism (VTE) (post-surgery)	$8,000	Spyropoulos AC, Lin J. Direct medical costs of venous thromboembolism and subsequent hospital readmission rates: an administrative claims analysis from 30 managed care organizations. J Manag Care Pharm. 2007 Jul-Aug;13(6):475-86. http://www.ncbi.nlm.nih.gov/pubmed/17672809 Maynard G, Stein J. Preventing hospital-aquired venous thromboembolism: A guide for effective quality improvement. Prepared by the Society of Hospital Medicine. AHRQ Publication No. 08-0075. Rockville, MD: Agency for Healthcare Research and Quality. August 2008. http://www.ahrq.gov/qual/vtguide/

Table A-3—Sources for Cost per Event Used in AHRQ Estimate of Cost Savings		
Condition	**Cost per Adverse Event**	**Source**
Pressure Ulcer	$17,000	Federal Register: April 30, 2008 (Volume 73, Number 84). Centers for Medicare and Medicaid Services. Medicare Program: Proposed Changes to the Hospital Inpatient Prospective Payment Systems and Fiscal Year 2009 Rates, 23528–23938 [08–1135]. http://www.ncbi.nlm.nih.gov/pubmed/19827228
Surgical Site Infection (SSI)	$21,000	CDC (Scott, RD), The Direct Medical Costs of Healthcare-Asssociated Infections in U.S. Hospital and the Benefits of Prevention. March 2009. Available at http://www.cdc.gov/ncidod/dhqp/pdf/Scott_CostPaper.pdf
Ventilator-Associated Pneumonia	$21,000	CDC (Scott, RD), The Direct Medical Costs of Healthcare-Asssociated Infections in U.S. Hospital and the Benefits of Prevention. March 2009. Available at http://www.cdc.gov/ncidod/dhqp/pdf/Scott_CostPaper.pdf
Catheter-Associated Urinary Tract Infection	$1,000	CDC (Scott, RD), The Direct Medical Costs of Healthcare-Asssociated Infections in U.S. Hospital and the Benefits of Prevention. March 2009. Available at http://www.cdc.gov/ncidod/dhqp/pdf/Scott_CostPaper.pdf
Adverse Drug Event	$5,000	Bates DW, Cullen DJ, Laird N, et al. Incidence of adverse drug events and potential adverse drug events. JAMA 1995;274:29-34. http://www.ncbi.nlm.nih.gov/pubmed/7791255
Obstetrical (OB) Adverse Event	$3,000	AHRQ Researcher Stanley Davis
Injury from Fall	$7,234	Federal Register: April 30, 2008 (Volume 73, Number 84). Centers for Medicare and Medicaid Services. Medicare Program: Proposed Changes to the Hospital Inpatient Prospective Payment Systems and Fiscal Year 2009 Rates, 23528–23938 [08–1135]. http://www.ncbi.nlm.nih.gov/pubmed/19827228

Table A-3—Sources for Cost per Event Used in AHRQ Estimate of Cost Savings		
Condition	**Cost per Adverse Event**	**Source**
All Other HACs	$17,000	HHS computation based on costs above

Source: See column 3, above.

Table A-4—AHRQ's Calculation of Cost Savings for 2011					
PfP Hospital Acquired Condition	**Estimated Cost per HAC**	**2010 Count of HACs (Rounded)**	**2011 Normalized Count of HACs (Rounded)**	**Reduction in HACs (2010 to 2011) (Rounded)**	**Estimated Cost Savings (2010-2011) (Not-Rounded)**
ADE	$5,000	1,621,000	1,594,000	27,000	$135,000,000
CAUTI	$1,000	400,000	370,000	30,000	$30,000,000
CLABSI	$17,000	18,000	17,000	1,000	$17,000,000
Falls	$7,234	260,000	260,000	0	0
OB-Adverse Events	$3,000	82,000	82,000	0	0
Pressure Ulcers	$17,000	1,320,000	1,320,000	0	0
SSI	$21,000	96,000	82,000	14,000	$294,000,000
VAP	$21,000	38,000	35,000	3,000	$63,000,000
VTE	$8,000	16,000*	15,000	1,000	$8,000,000
All Other HACs	$17,000	894,000	875,000	19,000	$323,000,000
Totals		**4,745,000**	**4,650,000**	**95,000**	**$870,000,000**

Source: Noel Eldridge, "Partnership for Patients National Data for Measured Hospital-Acquired Conditions: Data for 2010 & 2011 and Preliminary Data for Q1 2012, AHRQ Draft for CMMI Review," June 7, 2103. Center for Quality Improvement and Patient Safety, AHRQ.

*2010-2012 VTE data will be corrected in Final 2012 data to include all diagnosed cases of PEs.

Table A-5—AHRQ's Calculation of Cost Savings for 2012					
PFP Hospital Acquired Condition	**PFP Cost per HAC**	**2010 Count of HACs (Rounded)**	**2012 Normalized Count of HACs (Rounded)**	**Reduction in HACs (2010 to 2012 Rounded)**	**Reduction in Costs (2010 to 2012 not rounded)**
ADE	$5,000	1,621,000	1,372,000	249,000	$1,245,000,000
CAUTI	$1,000	400,000	350,000	50,000	$50,000,000
CLABSI	$17,000	18,000	17,000	1,000	$17,000,000
Falls	$7,234	260,000	230,000	30,000	$217,020,000
OB-Adverse Events	$3,000	82,000	77,000	5,000	$15,000,000
Pressure Ulcers	$17,000	1,320,000	1,300,000	20,000	$340,000,000
SSI	$21,000	96,000	82,000	14,000	$294,000,000
VAP	$21,000	38,000	34,000	4,000	$84,000,000
VTE	$8,000	28,000	32,000	-4,000	-$32,000,000
All Other HACs	$17,000	894,000	843,000	51,000	$867,000,000
Totals		**4,757,000**	**4,337,000**	**420,000**	**$3,097,020,000**

Source: Noel Eldridge, "PfP's 'AHRQ National Scorecard' or National HAC Rate: Updated with Final 2012 Data. Updated June 2, 2014.". Center for Quality Improvement and Patient Safety, AHRQ.

Estimation of Cost Savings from Other Sources

The second method for estimating cost savings uses a different method, different estimates of cost per event, and different data from the AHRQ estimate. This second estimate was requested from the Evaluation Contractor because AHRQ inpatient chart data from the MPSMS are available only annually, whereas more frequent updates would better serve PfP management. All focus-area-specific estimates except the estimate for readmissions are therefore partial estimates limited to available data and cost estimates from the literature as follows: OB-EED estimates cover reduced costs due to estimated reduced use of the NICU only, and only for hospitals reporting the data to a HEN; VAP, Falls, and OB-Other estimates cover only hospitals reporting to their HEN or NDNQI, depending on the measure (no extrapolation); SSI covers SSI for only two of many relevant surgical conditions; CAUTI and CLABSI measures cover hospital units reporting to NHSN only; pressure ulcers covers only high-stage pressure ulcers; and VTE baseline data are for Q2-Q4 2011 rather than 2010, due to data issues in earlier data.

- Various sources of data were used to calculate the number of adverse events averted (the data sources are further described in Appendices B and C):
  - NHSN data were used to calculate averted SSI, CAUTI, and CLABSI events.
  - NDNQI data were used to calculate averted falls with injury events.
  - HEN-submitted data were used to calculate averted EEDs and OB-Other events.
  - NDNQI and HEN-submitted data were used to calculate VAP events.[A-13]
  - Medicare claims data processed by CMS' Health Policy and Data Analysis Group were used to calculate averted VTE and Pressure ulcer events and readmissions.
- 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 Agency for Healthcare Research and Quality (AHRQ) PSI indicators depends, and/or use of 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 submitted data are used.

[A-13] Represents estimated events and costs avoided prior to hospitals shifting to new measures developed by the CDC. Estimates from the new measures will be included as soon as these data become available.

- The basic calculation of numbers of adverse events averted (for a given time period of interest) is as follows:
    - Number of Adverse Events Averted = Number of Adverse Events Expected (that is, projected) – 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 data collected by NHSN on device-associated, health care-associated infections (CLABSI and CAUTI) also feature standardization of current patient populations to patient characteristics from a baseline period through the SIR quantity, as well as data on device utilization through the UR quantity. For CLABSI and CAUTI estimates, where NHSN data were used, changes over time in SIR and UR were also incorporated into the calculations.
- Where data were limited to a specific payer population (Medicare), the analysis used historical ratios of adverse events among Medicare patients to events among non-Medicare patients to create an all-payer averted events number. Additional details of all calculations, including the specific measures used and the cost per event drawn from the literature, were provided to PfP leadership in a separate report between May 2013 and July 2013.

Table A-6 shows the estimates for cost per event that were used, based on a literature review by the Evaluation Contractor in 2012-2013, and new analysis of data from the Hospital Cost and Utilization Project (HCUP) in 2014. A range of estimates are available from literature sources, so the following criteria were considered in selecting those shown below:

- Baselines closest to 2010.
- Datasets with larger numbers of HENs and hospitals.
- Estimates based on index admission only (rather than extending to time periods beyond the hospitalization).
- Estimates on the lower-end of the range of cost estimates.
- Estimates previously used by HHS or AHRQ for consistency.
- Micro-costing methods vs. hospital charges.
- Datasets that include multiple payers.

The HCUP analysis provided original cost-per-event estimates for Medicare and non-Medicare populations for pressure ulcers (PSI-03) and VTE (PSI-12), and all-populations measures for the three obstetrical harm measures other than early elective delivery (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 was available at the time of analysis for 12 states.

Table A-6—Estimated Cost Per Event		
Adverse Event Area	**Estimated Cost per Event**	**Source**
CAUTI (ICU and Wards)	$1,000	Scott II, R. Douglas. "The Direct Medical Costs of Healthcare-Associated Infections in U.S. Hospitals and the Benefits of Prevention." Atlanta, GA: Centers for Disease Control Coordinating Center for Infectious Diseases; 2009. Available at: http://www.cdc.gov/hai/pdfs/hai/scott_costpaper.pdf], Accessed April 12, 2013.
CLABSI (ICU and Wards)	$17,000	Srinivasan A., M. Wise, M. Bell, et al. "Vital Signs: Central Line-Associated Blood Stream Infections United States, 2001, 2008, and 2009." Morbidity and Mortality Weekly Report. 2011; 60(8): 243-248. Available at: http://www.cdc.gov/mmwr/pdf/wk/mm6008.pdf. Accessed Apr 24, 2013.
EED	$7,875 for EEDs resulting in NICU; NICU stays estimated to equal 0.0996 x EEDs	1. Bailit, Jennifer L., Kimberly D. Gregory, "Maternal and neonatal outcomes by labor onset type and gestational age." American Journal of Obstetrics and Gynecology, vol. 202, pp. 245.e1-12. [Average NICU LOS for all early deliveries at 37 to 38 weeks gestational age] 2. March of Dimes. "Special Care Nursery Admissions." http://www.marchofdimes.com/peristats/pdfdocs/nicu_summary_final.pdf. Accessed April 17, 2013 [Average charge for a NICU stay for babies 37 to 38 gestational weeks] 3. Anderson, Gerard F. "From 'Soak The Rich' To 'Soak The Poor': Recent Trends in Hospital Pricing."Health Affairs, vol. 26, No. 3, 2007, pp. 780-789. [Cost-to-charge ratio] 4. Ehrenthal, D. B., Hoffman, M. K., Jiang, X., and Ostrum, G. "Neonatal Outcomes After Implementation of Guidelines Limiting Elective Delivery Before 39 Weeks of Gestation." Obstetrics and Gynecology, vol. 118, no. 5, 2011, pp. 1047-1055 [Estimated fraction of EEDs that lead to a NICU stay] 5. Friedman, B., J. La Mare, R. Andrews, and D. McKenzie. "Practical options for estimating cost of hospital inpatient stays." Journal of Health Care Finance, 2002, vol. 29, no. 1, pp. 1-13. 6. Zhan, Chunliu and Marlene R. Miller. "Excess Length of Stay, Charges, and Mortality Attributable to Medical Injuries During Hospitalization." JAMA. Vol. 290 , No. 14, October 8, 2003, pp 1868-1874.
Falls with injury	$663	Coomer, N., K. Dalton, and A. Kandilov. "Analysis Report: Estimating the Incremental Costs of Hospital-Acquired Conditions (HACs)." Final Report. Research Triangle Park, NC: RTI International. April 2012.
PSI-03—PrUs	15,394 (Medicare) $40,500 (Non-Medicare)	Analysis by the Evaluation Contractor of the difference in Medicare and non-Medicare hospital costs during the index stay between those with and without an adverse event, using a matched sample from the 2009 to 2011 HCUP data that was available at the time of analysis for 12 states.

Table A-6—Estimated Cost Per Event		
Adverse Event Area	**Estimated Cost per Event**	**Source**
PSI-12—VTE	$14,189 (Medicare) $22,240 (Non-Medicare)	Analysis by the Evaluation Contractor of the difference in Medicare and Non-Medicare hospital costs during the index stay between those with and without an adverse event, using a matched sample from the 2009 to 2011 HCUP data that was available at the time of analysis for 12 states.
PSI-17—Injuries to Neonate	$920	Analysis by the Evaluation Contractor of the difference in all-payer hospital costs during the index stay between those with and without an adverse event, using a matched sample from the 2009 to 2011 HCUP data that was available at the time of analysis for 12 states.
PSI-18—Obstetric Trauma-Vaginal Delivery with Instrument	$92	Analysis by the Evaluation Contractor of the difference in all-payer hospital costs during the index stay between those with and without an adverse event, using a matched sample from the 2009 to 2011 HCUP data that was available at the time of analysis for 12 states.
PSI-19—Obstetric Trauma-Vaginal Delivery without Instrument	$158	Analysis by the Evaluation Contractor of the difference in all-payer hospital costs during the index stay between those with and without an adverse event, using a matched sample from the 2009 to 2011 HCUP data that was available at the time of analysis for 12 states
Readmissions	$8,808 (Non-Medicare) $10,100 (Medicare)	U.S. Dept. of Health and Human Services. “National Patient Safety Initiative, Explanation of Estimates—The Potential Effects of Reducing Harm During Hospitalization and Hospital Readmissions.” Internal HHS Working Draft. Mar, 2011
SSI (Colon and Hysterectomy)	$21,000	CDC "2011 National and State Healthcare-Associated Infections Standardized Infection Ratio Report: Using Data Reported to the National Healthcare Safety Network as of September 4, 2012."Available at: http://www.cdc.gov/hai/pdfs/SIR/SIR-Report_02_07_2013.pdf
VAP	$21,000	Scott II, R. Douglas. “The Direct Medical Costs of Healthcare-Associated Infections in U.S. Hospitals and the Benefits of Prevention.” Atlanta, GA: Centers for Disease Control Coordinating Center for Infectious Diseases; 2009. Available at: http://www.cdc.gov/hai/pdfs/hai/scott_costpaper.pdf], Accessed April 12, 2013.

Source: See column 3, above.

Table A-7 provides the results of the estimates (non-rounded).

Table A-7—National Estimated Events and Costs Averted, Using Data Available in February 2014

Adverse Event	Hospitals Nationally		Data Source	Baseline Period	Follow-up Period	Notes
	Number of Averted Events	Cost Savings				
PrU	112	$2,595,542	PSIs computed on Medicare claims data	Q2-Q4 2011	CY 2012 – Q3 2013	Extrapolated to a national number under assumption that non-Medicare outcomes relative to Medicare remained constant from 2010 onward
VTE	2,285	$40,626,508	PSIs computed on Medicare claims data	Q2-Q4 2011	CY 2012 – Q3 2013	Extrapolation from Medicare to non-Medicare as above
VAP[a]	1,138	$23,906,377	NDNQI data augmented with data from HEN monthly reports for 3 HENs	NDNQI: CY 2011 HEN baselines vary	NDNQI: CY 2012–Q2 2013 HEN follow-up periods vary	
Readmissions	293,072	$2,821,422,333	Medicare claims (from the CMS Data and Policy Group to the Evaluation Contractor)	CY 2010	Q1 2011 – October 2013	Extrapolated to a national number under assumption that non-Medicare outcomes relative to Medicare remained constant from 2008 onward
OB-EED	16,088	$12,618,944	HEN submitted data	Baseline periods vary by HEN	Follow-up periods vary by HEN	Averted EEDs are used to calculate averted neonatal intensive care unit (NICU) stays

Table A-7—National Estimated Events and Costs Averted, Using Data Available in February 2014						
Adverse Event	Hospitals Nationally		Data Source	Baseline Period	Follow-up Period	Notes
	Number of Averted Events	Cost Savings				
CLABSI	9,782	$166,287,056	NHSN data, include events from both ICUs and non ICUs	CY 2010	CY 2011 – Q2 2013	Baseline data (2010) is representative of the hospital population submitting data to NHSN prior to the CMS Inpatient Prospective Payment System (IPPS) reporting mandate of ICUs in Q1 2011
Falls with Injury	6,397	$4,241,146	NDNQI data	CY 2011	2012 – Q3 2013	
OB-Other	75 (PSI-17) 582 (PSI-18) 4,925 (PSI-19)	$68,657 (PSI-17) $53,557 (PSI-18) $778,129 (PSI-19)	HEN submitted data	Baseline periods vary by HEN	Follow-up periods vary by HEN	
CAUTI	-712	($711,820)	NHSN data, include events from both ICUs and non ICUs	CY 2010	2011 - Q2 2013	Baseline data (2010) is representative of the hospital population submitting data to NHSN prior to the CMS IPPS reporting mandate of ICUs in Q1 2012

Table A-7—National Estimated Events and Costs Averted, Using Data Available in February 2014						
Adverse Event	Hospitals Nationally		Data Source	Baseline Period	Follow-up Period	Notes
	Number of Averted Events	Cost Savings				
SSI[a]	1,808	$37,974,126	NHSN data, include events from both colon and abdominal hysterectomy SSIs	CY 2010	2011-Q2 2013	Baseline data (2010) is representative of the hospital population submitting data to NHSN prior to the CMS IPPS reporting mandate in Q1 2012
Total Estimated Cost Savings		**$3,109,860,555**				

Source: See column 4, above.

Notes: Negative numbers of averted events indicate that more adverse events occurred in the follow-up period than were expected from baseline rates. In these instances, estimated costs will be higher (shown as negative cost savings).

Adverse drug events (ADEs) currently are not calculated due to a lack of common measures of ADEs across HENs. See Appendix A for information on how averted events and costs were calculated.

[a]Represents estimated events and costs avoided prior to hospitals shifting to new measures developed by the Centers for Disease Control and Prevention (CDC). Estimates from the new measures will be included here as soon as these data become available.

APPENDIX B: DATA SOURCES

This appendix provides more detail about the HEN data and National Data Sources used in the report.

HEN Data

The Evaluation Contractor calculated rates from HEN data for seven measures included in this report:

- The Joint Commission [TJC] Perinatal Care (PC)-01 (slight variations in measure definition exist for three of the 21 reporting HENs)
- AHRQ PSI-12 (VTE)
- AHRQ PSI-03 (pressure ulcers)
- AHRQ PSI-17 (birth trauma)
- AHRQ PSI-18 (obstetrical trauma)
- AHRQ PSI-19 (obstetrical trauma)
- 30-day all-payer readmissions rate (slight variations exist among HENs, including whether the rate is the rate of readmissions to the reporting hospital or the rate of readmissions to any hospital)

Numerators and denominators for the baseline were summed to calculate an aggregate baseline rate across HENs, and the same aggregation was conducted for the most current period available in each HEN's data. The baseline and most current periods for each HEN were at least 3 months in length (often the baseline period is a year), and the number of hospitals in each HEN's baseline and current period are within 15 percent of each other to assure rough comparability in the reporting hospital population. Table B-1 shows the number of reporting HENs and hospitals whose data were included for each measure.

Table B-1—Number of HENs and Hospitals Included in Measures Calculated from HEN-Submitted Data							
Indicator	**TJC PC-01 (EED)**	**PSI-17 (Birth Trauma)**	**PSI-18 (OB Trauma-Instr.-Assisted Deliveries)**	**PSI-19 (OB Trauma–Deliveries w/out Instr.)**	**PSI-12 (Perioperative PE or DVT)**	**PSI-03 (Stage 3+ Pressure Ulcers)**	**30-Day All-Cause, All-Payer Readmission Rate**
Number of HENs submitting	21	23	23	22	21	12	18
Number of Hospitals in Baseline	1597	1199	1450	1489	1761	1164	2301
Number of Hospitals in Current Period	1540	1177	1392	1450	1717	1139	2205

Source: Data submitted by HENs in February 2014.

National Data Sources

Readmissions data from the Centers for Medicare & Medicaid Services (CMS)

- Thirty-day all-cause readmissions data are generated by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS, for Medicare fee-for-service (FFS) beneficiaries discharged from HEN-aligned and non-aligned hospitals, January 2010 through November 2013. Readmission rates are generated from claims for beneficiaries who were enrolled in FFS Medicare Part A during the month of the index admission and are limited to acute care hospitals. A readmission will also count as a new index admission.
- A Medicare claim is not final until a few months after it is first received and has undergone processing and adjudication. Thus, results obtained from non-final claims data may vary slightly from those obtained from final data; however, CMS has developed a model that uses non-final data to project the readmission rates (with associated confidence intervals) that will be seen in the final data, thus, data through November 2013 are included in the analysis.

NHSN Data

The CDC provides quarterly data from the NHSN. The data extend through Q2 2013, and include HEN-aligned hospitals (including those aligned with Indian Health Services [IHS]) and non-aligned hospitals.

The data periods vary by measure, corresponding to the periods when hospitals receiving inpatient prospective payment for Medicare services were required by Medicare to report NHSN measures for their intensive care unit (ICU). Given hospitals' requirement to report, the NHSN data provide a strong representation of program progress on the measures for the periods available, for PfP as a

whole and for most HENs.[B-1] Since critical access hospitals (CAHs) were excluded from the Medicare NHSN reporting requirement, the data for these hospitals are less complete. The data periods are:

- 2011 through Q2 2013 for central line-associated blood stream infection (CLABSI) standardized infection ratio (SIR).
- Q1 2011 through Q2 2013 for central line utilization ratio (UR) (PfP-aligned hospitals only).
- Q1 2012 through Q2 2013 for facility-wide surgical site infection (SSI)-colon surgery and SSI-abdominal hysterectomy SIRs.
- Q1 2012 through Q2 2013 for CAUTI SIR.
- Q1 2012 through Q2 2013 for catheter UR (PfP-aligned hospitals only).

NDNQI Data

The NDNQI is an American Nursing Association (ANA) database that is housed at and administered by the University Of Kansas School Of Nursing. As of November 2013, 1,941 hospitals within the 50 states and the District of Columbia were members of the NDNQI.[B-2,B-3] Hospitals paying a membership fee to NDNQI submit information on nursing-sensitive process, outcome, and structural measures at the hospital unit level on a quarterly basis.[B-4,B-5] Data through Q3 2013 is included in this report.

- Data received from NDNQI include catheter-associated urinary tract infection (CAUTI), CLABSI, falls, hospital-acquired pressure ulcers (HAPU), and ventilator-associated pneumonia (VAP) rates. All measures use a baseline of calendar year 2011 and a follow up period of Q3 2013. To ensure that the trends represent a real change in the measure among reporting hospitals, rather than a change in the mix of hospitals, data include only hospitals reporting both in the current period (Q3 2013) and in at least 80 percent of the nine previous quarters (Q1 2011 through Q2 2013).
- Favorable trends among NDNQI-reporting hospitals likely overstate the success achieved nationally, since it is likely that hospitals that have been willing to pay to participate in NDNQI since 2011 are achieving better results than the average hospital not participating in the NDNQI.

B-1 In addition, many states also require hospitals to report healthcare associated infections, typically through the NHSN system.
B-2 NDNQI data were supplied by the ANA. The ANA disclaims responsibility for any analyses, interpretations or conclusions.
B-3 Hospital numbers by state are located here: http://www.nursingworld.org/ndnqi2.
B-4 Hospital numbers by state are located here: http://www.nursingworld.org/ndnqi2.
B-5 Information on quarterly data submission can be found: https://www.nursingquality.org/Datasubmission.aspx.

Medicare Claims Data: AHRQ PSIs

PSI measures for CLABSI (PSI-07), PrUs (PSI-03), and VTE (PSI-12) are generated by the Health Policy and Data Analysis Group in the Office of Enterprise Management at CMS, for Medicare fee-for-service (FFS) beneficiaries discharged from HEN-aligned and non-aligned hospitals, through Q3 2013.

All national hospitals (e.g., IPPS hospitals, CAH, Maryland and Puerto Rico hospitals, and cancer hospitals) are included in these data; however, the data were restricted to those hospitals with adequate POA reporting. A hospital's data for a quarter were excluded if more than 5 percent of the hospital's diagnoses that were not exempt from reporting POA codes had inappropriate POA values.[B-6,B-7]

The baseline period for the PSI measures excludes Q1 2011. There were problems in that quarter with miscoding of POA indicators, which compromised the integrity of the PSI rates. In Q1 2011, the number of diagnosis codes that IPPS hospitals were required to report changed from 9 to 25; without special adjustment the data prior to Q2 2011 are non-comparable.

B-6 Every year CMS publishes a list of diagnoses that are exempt from reporting POA codes.

B-7 Inappropriate POA values included blank, invalid, or those wrongly indicating that the diagnosis is exempt from reporting POA values.

APPENDIX C: SUMMARY OF MEASURE LIMITATIONS

Table C-1 provides a summary of the limitations of each of the measures used in the report.

Table C-1—Limitations of Measures Used[a]

AEA–Measure Name (Source)	Note Here if Only a Subset of Care in the Area is Measured	Length of Trend: Number of Quarters Covered	Claims-Based Measure, Likely to Under-Count Harms	Note if Medicare Only (Otherwise All-Payer)	Completeness: Number of Hospitals Included, of 5,196 U.S. Short-Term Acute Care Hospitals (Current Period)	Voluntary Subset of Hospitals Subject to Potential Bias?	Stronger Measures for the AEA, Where Multiple Measures Exist	Why Stronger?
CAUTI - CAUTI per 1,000 Catheter Days, All Tracked Units (NDNQI)		11			717	Yes		
CAUTI – QIO provided results		5 months pre/post			148	Yes		
CAUTI - CAUTI SIR (Observed/Expected) (ICUs) (NHSN)	ICU only	6			3,175		Stronger	More complete; mandatory for PPS hospitals
CAUTI - CAUTI SIR (Non-ICU Units) (NHSN)	non-ICU only	6			1,346	Yes		
CLABSI - CLABSI per 1,000 Central Line Days, All Tracked Units (NDNQI)		11			809	Yes		

Table C-1—Limitations of Measures Used[a]								
AEA–Measure Name (Source)	**Note Here if Only a Subset of Care in the Area is Measured**	**Length of Trend: Number of Quarters Covered**	**Claims-Based Measure, Likely to Under-Count Harms**	**Note if Medicare Only (Otherwise All-Payer)**	**Completeness: Number of Hospitals Included, of 5,196 U.S. Short-Term Acute Care Hospitals (Current Period)**	**Voluntary Subset of Hospitals Subject to Potential Bias?**	**Stronger Measures for the AEA, Where Multiple Measures Exist**	**Why Stronger?**
CLABSI - CLABSI SIR, ICUs (NHSN)	ICU only	10			3,163		Stronger	More complete; mandatory for PPS hospitals
CLABSI - CLABSI SIR, Non-ICUs (NHSN)	non-ICU only	10			1,256	Yes		
CLABSI – QIO provided results		5 months pre/post			667	Yes		
CLABSI - CR-BSI per 1,000 Discharges (AHRQ PSI-07) (Medicare Claims)	CRBSI narrower diagnosis than CLABSI	10	Yes	Medicare				
Falls - Falls per 1,000 Patient Days (NDNQI)		11			1,340	Yes		
Falls - Falls With Injury per 1,000 Patient Days (NDNQI)		11			1,340	Yes	Stronger	More directly associated with harms, broader than hip fracture measure

Table C-1—Limitations of Measures Used[a]								
AEA–Measure Name (Source)	**Note Here if Only a Subset of Care in the Area is Measured**	**Length of Trend: Number of Quarters Covered**	**Claims-Based Measure, Likely to Under-Count Harms**	**Note if Medicare Only (Otherwise All-Payer)**	**Completeness: Number of Hospitals Included, of 5,196 U.S. Short-Term Acute Care Hospitals (Current Period)**	**Voluntary Subset of Hospitals Subject to Potential Bias?**	**Stronger Measures for the AEA, Where Multiple Measures Exist**	**Why Stronger?**
Falls - Post-Operative Hip Fracture per 1,000 Discharges (AHRQ PSI-08) (Medicare Claims)	Small subset of falls result in hip fracture	10	Yes	Medicare				
Pressure Ulcers - Patients with Hospital-Acquired PrU, Stages 2+, per 1,000 Discharges (NDNQI)		11			1,341	Yes	Equal	
Pressure Ulcers - PrU per 1,000 Discharges (Stages 3+) (AHRQ PSI-03) (Medicare Claims)	Most severe PrU	10	Yes	Medicare			Equal	
Pressure Ulcers - PrU per 1,000 Discharges (Stages 3+) (AHRQ PSI-03) (HENs)	Most severe PrU	Mixed	Yes		1,194	Yes		

Table C-1—Limitations of Measures Used[a]								
AEA–Measure Name (Source)	**Note Here if Only a Subset of Care in the Area is Measured**	**Length of Trend: Number of Quarters Covered**	**Claims-Based Measure, Likely to Under-Count Harms**	**Note if Medicare Only (Otherwise All-Payer)**	**Completeness: Number of Hospitals Included, of 5,196 U.S. Short-Term Acute Care Hospitals (Current Period)**	**Voluntary Subset of Hospitals Subject to Potential Bias?**	**Stronger Measures for the AEA, Where Multiple Measures Exist**	**Why Stronger?**
Readmissions - Medicare FFS 30-Day All-Cause Readmissions (Medicare Claims)		15	Yes	Medicare			Stronger	Nearly complete for Medicare so not subject to reporting bias; measure exactly same across hospitals
Readmissions - 30-Day All-Cause Readmissions (HENs)		Mixed			2,205	Yes		
SSI – SSI - Colon Surgery SIR (NHSN)	SSI for one procedure	6			3,331		Equal	
SSI – SSI - Abdominal Hysterectomy SIR (NHSN)	SSI for one procedure	6			3,326		Equal	
VAE - VAP per 1,000 Ventilator Days (NDNQI)[b]		11			547	Yes		

Table C-1—Limitations of Measures Used[a]								
AEA–Measure Name (Source)	**Note Here if Only a Subset of Care in the Area is Measured**	**Length of Trend: Number of Quarters Covered**	**Claims-Based Measure, Likely to Under-Count Harms**	**Note if Medicare Only (Otherwise All-Payer)**	**Completeness: Number of Hospitals Included, of 5,196 U.S. Short-Term Acute Care Hospitals (Current Period)**	**Voluntary Subset of Hospitals Subject to Potential Bias?**	**Stronger Measures for the AEA, Where Multiple Measures Exist**	**Why Stronger?**
VTE - Perioperative PE or DVT per 1,000 Surgical Discharges (AHRQ PSI-12) (Medicare Claims)	Only peri-operative VTE		Yes	Medicare			Stronger	Nearly complete for Medicare, so not subject to reporting bias; measure exactly same across hospitals
VTE - Perioperative PE or DVT per 1,000 Surgical Discharges (AHRQ PSI-12) (HENs)	Only perioperative VTE	Mixed			1,717	Yes		
OB-EED - Early Elective Delivery Rate (TJC PC-01) (HENs)		Mixed			1,540	Yes		
OB-Other - Injury to Neonate (AHRQ PSI-17) (HENs)	Subset of obstetric harms	Mixed	Yes		1,177	Yes	Equal	
OB-Other - Obstetrical Trauma (AHRQ PSI-18) (HENs)	Subset of obstetric harms	Mixed	Yes		1,392	Yes	Equal	

Table C-1—Limitations of Measures Used[a]								
AEA–Measure Name (Source)	Note Here if Only a Subset of Care in the Area is Measured	Length of Trend: Number of Quarters Covered	Claims-Based Measure, Likely to Under-Count Harms	Note if Medicare Only (Otherwise All-Payer)	Completeness: Number of Hospitals Included, of 5,196 U.S. Short-Term Acute Care Hospitals (Current Period)	Voluntary Subset of Hospitals Subject to Potential Bias?	Stronger Measures for the AEA, Where Multiple Measures Exist	Why Stronger?
OB-Other - Obstetrical Trauma (AHRQ PSI-19) (HENs)	Subset of obstetric harms	Mixed	Yes		1,450	Yes	Equal	
CAUTI - CAUTI per 1,000 Catheter Days, All Tracked Units (NDNQI)		11			717	Yes		
CAUTI - CAUTI SIR (Observed/Expected) (ICUs) (NHSN)	ICU only	6			3,175		Stronger	More complete; mandatory for PPS hospitals
CAUTI - CAUTI SIR (Non-ICU Units) (NHSN)	non-ICU only	6			1,346	Yes		

[a] Measures from the AHRQ National Scorecard are not included in this table; more information on those measures can be obtained by contacting Noel.Eldridge@AHRQ.hhs.gov.
[b] Concerns about the definition of VAP used in this measure resulted in a change in the CDC's definition, however, data for the new definition are not yet available.

APPENDIX D: AHA/HRET INITIAL ANALYSES OF ENGAGEMENT AND OUTCOMES

The Evaluation Contractor did not conduct this analysis. The analysis was conducted by the AHA/HRET HEN.

Are Hospitals More Engaged in Quality Improvement Efforts More Successful? An Assessment of HRET HEN Hospitals over 26 Months Executive Summary—April 28, 2014

Overview

Hospitals participating in the national HEN initiative have shown considerable improvements in areas the HEN initiative has targeted. HRET has performed a series of analyses designed to explore the relationship between a variety of engagement measures and HEN outcomes. While the analyses cannot establish causality, they do point to a clear relationship between several measures of higher engagement and superior performance on HEN outcomes.

Methods

Population. HRET works with 1512 hospitals drawn from 31 states. Data used in the following analyses are from the 1356 acute, critical access, and children's hospitals participating in the HRET HEN as of February 24, 2014.

Measuring Engagement. Engagement in the HEN project has multiple dimensions that no single measure can adequately capture. HRET identified four general categories of engagement measures. These included:

1. The extent to which hospitals participated in HEN activities sponsored by HRET and its project partners
   a. Participation in HRET HEN events
   b. Participation in the Improvement Leader Fellowship
2. The extent to which hospitals implemented recommended HEN activities and processes within their organization
   a. Leadership engagement
   b. Patient and family engagement
   c. "Eliminating Harm Across the Board" storyboards
3. The breadth of HEN-targeted areas the hospital worked to measure and improve
   a. Number of targeted topics
   b. Data submission

4. How the hospital's level of engagement was perceived by the HRET project partner working most closely with them
   a. Partner assessment of hospital engagement
   b. Partner assessment of leadership engagement
   c. Resistance to greater engagement

Measuring Outcomes. We assessed two outcome measures. The primary outcome measure was the CMS composite measure used to assess the percent of eligible topics on which a hospital demonstrated success. Hospitals working on fewer topics and submitting less data would inevitably have lower scores on this composite measure. To avoid confounding in some of our analyses, we constructed a second outcome measure that is limited to progress on the topics for which the hospital is reporting data.

Analyses. Models were tested that included each of the measures within the four engagement categories described above. We tested for the overall significance of the model and examined the total variance accounted for. We also examined the significance of each predictor within the four models.

Results

1. The extent to which hospitals participated in HEN activities sponsored by HRET and its project partners

In a model including the two activity participation measures, both participation in HEN events and the level of a hospital's involvement in the Improvement Leader Fellowship program were significant predictors of our primary outcome measure ($p<.01$). Hospitals sponsoring a Champion Improvement Leader Fellow met CMS goals on 13% more topics than hospitals not involved in the Fellowship program.

2. The extent to which hospitals implemented recommended HEN activities and processes within their organization

The overall model testing the relationship between the three hospital implementation measures and our primary outcome was significant at $p<.001$. Leadership engagement and the use of storyboarding were significant individual predicators at $p<.001$ while patient and family engagement was a marginally significant predictor ($p=.053$). Hospitals with high or very high scores on the patient engagement level had outcome scores 10% higher than hospitals with low patient engagement scores, and hospitals that have reported storyboards to track patient harms in their hospital had 15% higher composite measure scores than hospitals that have not implemented this storyboarding.

3. The breadth of HEN-targeted areas the hospital worked to measure and improve

For these analyses, we assessed the relationship between number of topics and amount of data submitted and a measure of how successful hospitals were on the topics for which they were collecting and reporting data. The model that included both the number of topics hospitals were

working on and the number of data points submitted was highly significant ($p<.001$) and accounted for 21% of the variance in the average amount of improvement observed on measures for which the hospital was reporting data. Both variables were significant independent predictors at $p<.01$. Hospitals working on the highest number of topics were significantly more likely to have met the outcome goals on those topics than hospitals working on the fewest number of topics. Moreover, hospitals reporting more data showed more improvement on targeted topics than hospitals reporting less data.

4. How the hospital's level of engagement was perceived by the HRET project partner working most closely with them

State hospital associations (SHA) provided their assessments of the extent to which hospitals were fully engaged in HEN, the extent to which hospital leaders were committed to the goals of Partnership for Patients, and hospitals' reluctance to more fully engage. The model including all three of these predictors was significant at $p<.001$ but only accounted for 3% of the variance in the percent of topics meeting CMS HEN goals. Hospitals perceived to have the highest levels of overall engagement and leadership engagement had significantly higher outcome scores ($p<.02$). Resistance to more extensive engagement was not a significant predictor. Hospitals perceived as most engaged and with leadership who were most engaged had 7% or higher scores on the outcome measure.

	Predictor	Outcome Tested	Result
1. The extent to which hospitals participated in HEN activities sponsored by HRET and its project partners	Participation in HRET HEN Events	1: Percent of Eligible Topics that have met the CMS Goal	Predictor significant ($p<0.01$)
	Improvement Leader Fellowship Participation	1: Percent of Eligible Topics that have met the CMS Goal	Predictor significant ($p<0.01$)
2. The extent to which hospitals implemented recommended HEN activities and processes within their organization	Leadership Engagement	1: Percent of Eligible Topics that have met the CMS Goal	Predictor significant ($p<0.001$)
	Patient and Family Engagement	1: Percent of Eligible Topics that have met the CMS Goal	Predictor significant ($p<0.053$)
	Eliminating Harm Across the Board Storyboards	1: Percent of Eligible Topics that have met the CMS Goal	Predictor significant ($p<0.001$)
3. The breadth of HEN-targeted areas the hospital worked to measure and improve	Number of Topics Hospitals are Engaged in with the HEN	2: Mean Improvement Level	Predictor significant ($p<0.01$)
	Data Submission	2: Mean Improvement Level	Predictor significant ($p<0.01$)
4. How the hospital's level of engagement was perceived by the HRET project partner working most closely with them	Partner Assessment of Engagement of Hospital Engagement	1: Percent of Eligible Topics that have met the CMS Goal	Predictor significant ($p<0.02$)
	Partner Assessment of Leadership Engagement	1: Percent of Eligible Topics that have met the CMS Goal	Predictor significant ($p<0.02$)
	Partner Assessment of Hospital Resistance to Greater Engagement	1: Percent of Eligible Topics that have met the CMS Goal	Not Significant

Limitations

HRET fully recognizes that none of this data provides clear evidence that more extensive engagement in HEN activity caused greater levels of improvement. We also acknowledge that the measures for both engagement and the outcomes we used all have limitations and could be further refined.

Summary

Across all four engagement domains we examined and all ten of the engagement measures we constructed, higher levels of engagement were associated with higher levels of attainment on the outcome measures targeted by HEN. In a combined analysis using all significant predictors and the outcome measure representing the average amount of improvement observed on topics hospitals were working on, the nine engagement measures accounted for 24% of the variance in this measure. While further analyses would enhance our understanding of this relationship, the reported results do support the claim that greater engagement in the HEN and HEN activities and priorities is related to higher performance and are fully consistent with the hypothesis that more extensive participation in HEN activities leads to better outcomes for patients.

Are Hospitals More Engaged in Quality Improvement Efforts More Successful? An Assessment of HRET HEN Hospitals over 26 Months—April 25, 2014

Overview

Hospitals participating in the national Hospital Engagement Network (HEN) initiative have shown considerable improvements in targeted areas. But because HEN is not a controlled experiment, efforts to show that HEN activities have caused these improvements are challenging. One approach to examining the feasibility of a potential causal relationship between HEN activities and successful outcomes is to assess the extent to which HEN hospitals engaged in activities and processes the HRET HEN was promoting. If hospitals more extensively engaged in our HEN had higher levels of success on the targeted HEN outcomes, a causal relationship is more probable. This document reports the results of a series of analyses designed to explore the relationship between a variety of engagement measures and HEN outcomes. We first describe four general categories of engagement we examined, along with the specific measures within each category. We then describe the two outcome measures we constructed. The document concludes by summarizing the results of the analyses. Collectively, they suggest a strong relationship between higher levels of engagement and better HEN outcomes.

Methods

Target Population

As the largest HEN, HRET works with 1512 hospitals drawn from 31 states. The analyses that follow use data collected during the first 26 months of the project (the HEN two-year base period plus the first two months of the option year). We included data from the 1356 acute, critical access, and children's hospitals participating in the HRET HEN as of February 24, 2014

Measuring Engagement

Engagement in the HEN project has multiple dimensions that no single measure can adequately capture. HRET began by identifying four general categories of engagement measures. These included:

1. The extent to which hospitals participated in HEN activities sponsored by HRET and its project partners
2. The extent to which hospitals implemented recommended HEN activities and processes within their organization
3. The breadth of HEN-targeted areas the hospital worked to measure and improve
4. How the hospital's level of engagement was perceived by the HRET project partner working most closely with them

The remainder of this section describes the specific measures we constructed to assess engagement within each of these four categories.

Activity Participation Measures

One plausible way to assess engagement is by examining the extent to which hospitals participated in HEN activities. Hospitals participating in more activities reflect higher levels of engagement. We constructed one measure of participation across all of the HEN activities HRET sponsored and a second measure based on participation in the Improvement Leader Fellowship program that was a central part of HRET's improvement strategy.

Participation in HEN EventsThe HRET HEN hosts many educational events to provide tools and resources to hospitals that they can use to achieve the goals of the Partnership for Patients (PfP). There have been over 9,300 hospital participants across all of our educational offerings from March 2012 through February 2014. Educational offerings include in-person and/or virtual onboarding sessions, topic-specific boot camps and webinars, the Improvement Leader Fellowship program, and focused affinity group meetings. Meetings conducted by our state hospital association (SHA) partners were not included because participant lists from those events were not readily accessible. In order to assess each hospital's participation in HRET HEN events, we counted the number of unique participants that have participated in at least one event from each hospital. Hospitals were placed into one of five groups based on the number of unique participants involved in at least one event.

Participation in Improvement Leader Fellowship. A major HRET HEN initiative was to support an Improvement Leader Fellowship program. This program provided extensive training in quality improvement, the science of rapid-cycle improvement, reliability, teamwork, patient and family engagement, safety culture, and guidance on how to lead improvement efforts to eliminate harm across the board. Some hospitals did not choose to sponsor an Improvement Leader Fellow; others had one or more Fellows, and some HRET HEN hospitals invested in one or more Fellows that demonstrated substantial engagement and capacity to lead improvement efforts in their hospital. We used this information to create an engagement measure with three categories (low=hospital had no participation in our Improvement Leader Fellowship program; medium=hospital had one or more Fellows; and high=hospital had one or more Champion Fellows). Champion Fellows are defined as HEN improvement leaders who have extensive quality and process improvement experience, manage a portfolio of projects, and mentor others Fellows. Sponsoring a Champion Fellow represents a high level of commitment and engagement of the hospital in the HRET HEN.

Hospital Implementation of Recommended Practices

A second way to assess engagement is by assessing the extent to which hospitals implemented activities and processes that the HRET HEN promoted. Three key implementation areas were identified for which data were available to construct engagement measures. These included how extensively leadership was engaged in promoting Partnership for Patients improvement efforts, how extensively hospitals implemented recommended processes for better engaging with patients and families, and whether the hospital used storyboarding that HRET promoted to call attention to patient harms.

Leadership Engagement. HRET invested considerable effort in encouraging hospital leaders to aggressively support the HEN-related improvement efforts in their hospital. Hospitals were encouraged to hold regular quality reviews aligned with the PfP goals, to publicly report HEN

quality data, to reinforce the message that all staff play a role in protecting patient safety, and to have a Board-level committee that regularly reviews quality data. We asked our HEN hospitals to report on how many of these activities they had in place. Based on the March 9, 2014 Level of Participation Report, we used this information to create a leadership engagement measure with three categories (low=0-1 of these activities taking place; medium=2-3 of these activities taking place; and high=all four of these activities taking place).

Patient and Family Engagement. HRET is convinced that effectively engaging patients and families is central to fostering a culture of safety and will lead to superior outcomes. The HRET HEN invested considerable resources promoting patient and family engagement with participating hospitals. Hospitals were encouraged to create planning checklists for patients prior to their admission, conduct shift-change huddles that include patients and families when possible, create a position or department that fosters and oversees patient engagement activities, include patients and families in one or more hospital-level committees, and have one or more patients who function as patient representatives on a hospital leadership or governing board. We asked our HEN hospitals to report on how many of these activities they had in place. Based on the March 9, 2014 Level of Participation Report, we used this information to create a third engagement measure with five categories (very low=0 of these activities taking place; low=1 activity taking place; medium=2 of these activities taking place; high=3 of these activities taking place; and very high=4-5 of these activities taking place).

"Eliminating Harm Across the Board" Storyboards. HRET has also encouraged our HEN hospitals to engage in a storyboarding process that tracks and displays all harms occurring in the hospital based on the CMS Model. This activity requires a considerable investment of time and a substantial commitment to transparency relating harms that occur at all levels of the organization. HRET obtained information from hospitals about whether they have implemented this process and have submitted a storyboard to HRET. We used this information to classify hospitals as either more engaged (have implemented and submitted a storyboard) or less engaged (have not submitted a storyboard).

Breadth of HEN Involvement

A third approach to assessing engagement is by examining the extent to which hospitals worked on, and collected and reported data for, as many of the HEN topics as were relevant to them. HRET constructed two specific measures reflecting this form of engagement: the percent of relevant HEN topics the hospital chose to work on and the amount of data hospitals submitted to HRET.

Number of Targeted Topics. HRET encouraged our HEN hospitals to work on all topics. Some hospitals worked on a large majority of the topics while others worked on relatively few. We regarded this as a final indicator of engagement. After excluding topics not applicable for the hospital (i.e. EED and OB adverse events in hospitals that do not deliver babies), we classified hospitals on each relevant topic on a three-point scale: (1=not working on the topic; 2=working on the topic but not submitting data; and 3=working on the topic and submitting data). We averaged these scores across all 11 topics and then classified our hospitals into thirds based on their level of engagement (low: engaged on fewest number of topics=average 2.5 or below; medium: engaged on

average number of topics=average 2.51-2.75; and high: engaged on higher number of topics=average 2.76-3.0).

Data Submission. Over the course of the project, much of our efforts have focused on the data: encouraging hospitals to collect and submit data across all 11 topics and analyzing and leveraging the data to support improvement projects. We summed the total number of data points (both process and outcome measures) that each hospital has submitted to HRET during the period of time covered by our analyses. Hospitals were placed into one of five groups based on the number of submitted data points (very low=10-200 data points; low=201-300 data points; medium=301-400 data points; high=401-600 data points; and very high=601 or more data points). Hospitals that did not submit sufficient data to reliably measure any outcomes were excluded from the analysis.

Perceptions of Engagement by HRET Project Partners (States)

A fourth approach to assessing engagement is by asking the SHA partners that work directly with all of the hospitals in the HRET HEN to provide their assessment of how extensively the hospital has been engaged in HEN activities. We utilized the standard CMS question and constructed three specific measures: an assessment of the extent to which the hospital was engaged; a measure of whether the leadership was engaged; and a final measure of whether the hospital resisted efforts to become more engaged in HEN activities.

State Assessment of Hospital Engagement. HRET partners closely with the SHAs to provide the education, tools, and resources that hospitals need to meet the outcome goals of the HEN initiative. Hospitals have differing levels of engagement in these activities, their overall participation, and their willingness to commit to the HEN project. To obtain an additional measure of hospital engagement, in April 2014, HRET asked each SHA partner to rate their HEN hospitals based on the hospital's level of engagement during the project's first two years. SHAs classified each hospital as high, moderate, or low based on their assessment of whether the hospital demonstrated close engagement with PfP.

State Assessment of Leadership Engagement. HRET also asked our SHA partners to rate each HEN hospital from their state based on their perception of the hospital's leadership. Hospitals were classified as high, moderate, or low in possessing strong leadership engagement with the HEN goals.

Resistance to Greater Engagement. In addition, HRET compiled SHA responses to a question about whether each hospital from their state was "mostly aloof from PfP despite efforts to engage them" more fully. Responses to this question were used to create a third perceptual measure of hospital engagement.

Measuring Outcomes

We assessed two outcome measures. The primary outcome measure was the CMS composite measure used to assess the percent of eligible topics on which a hospital demonstrated success. Hospitals working on fewer topics and submitting less data would inevitably have lower scores on this composite measure. To avoid confounding in some of our analyses, we constructed a second outcome measure that is limited to progress on the topics for which the hospital is reporting data.

Primary Outcome Measure. The first outcome measure was a composite outcome measure that represented the percentage of topics a hospital was eligible to work on that met the CMS-established goal. The composite outcome measure is calculated by taking the number of topics for which the hospital has met the CMS-established goal divided by the number of topics for which the hospital is eligible. The CMS-established goal could be met in three ways: 1) by demonstrating the required level of improvement (percent reduction); 2) demonstrating a sustained zero rate on the measure; or 3) meeting the benchmark rate. Benchmark rates were prescribed by CMS for measures listed in the PfP PEC: Hospital List Scoring Criteria document from March 2013. If a benchmark rate was not provided by CMS, benchmarks were calculated by HRET using CY2012 data using the methodology described in the PfP PEC: Hospital List Scoring Criteria document. Hospital rates on this measure could range from 0% (failed to meet CMS standard on any of the topics they were eligible to work on) to 100% (met the CMS success standard on all topics they were eligible to work on). For example, a hospital that was eligible for all 11 topics, but only submitting data and reaching the PfP goal for 5 topics would be at a 45 percent. A hospital that was eligible for 9 topics (a non-birthing hospital) and submitted data and reached the PfP goal for all 9 topics would be at 100 percent. Based on available data, scores ranged from 0% to 91%, with a mean of 42% (N=1,352, SD=22%).

Secondary Outcome Measure. As a secondary outcome measure, we classified each hospital's results on each of the 11 HEN-targeted outcomes into one of four categories (0=reporting data, no improvement; 1=showing 20% improvement on an outcome or 10% improvement in readmission; 2=meeting the CMS established goal; or 3=meeting the CMS established goal and showing broad engagement in related AEA efforts). Hospitals that had not submitted data were excluded from this analysis since improvement could not be assessed for them. Each hospital's rate equaled the average of all scores on topics for which they were submitting data. Based on available data, the average of the categorized improvement across applicable topics ranged from 0 to 3, with a mean of 1.56 and standard deviation of 0.49; those averaging a 0 showed no improvement on any topics for which they reported data, and those with an average of 3 met the CMS established goal and showed broad engagement in related Adverse Event Area efforts on all topics for which data had been submitted (N=1,254).

Analyses

Models were tested that included each of the measures within the four engagement categories described above. We tested for the overall significance of the model and examined the total variance accounted for. We also examined the significance of each predictor within the four models. Results of those analyses are reported in the section below.

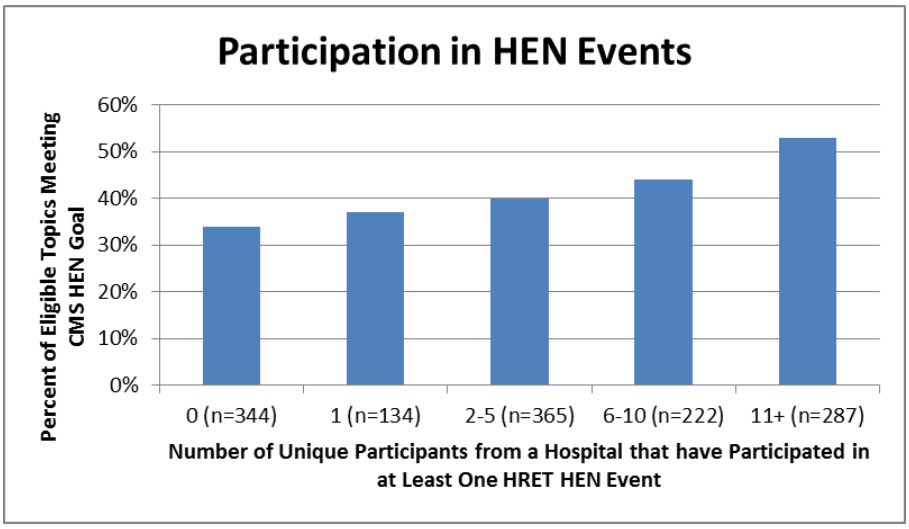
Results

Activity Participation Measures

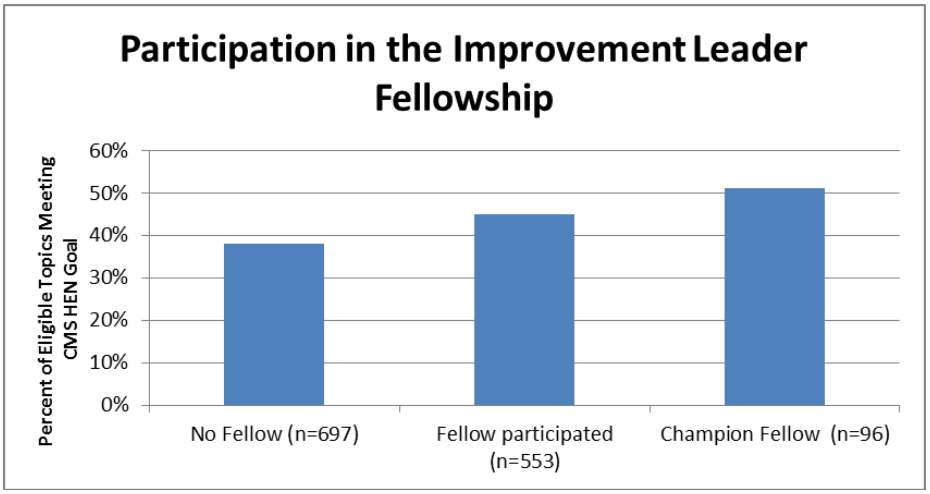
In a model including both activity participation measures, both participation in HEN events and the level of a hospital's involvement in the Improvement Leader Fellowship program were significant predictors of our primary outcome measure ($p<.01$). Together, these two variables accounted for 10% of the variance in the composite outcome score. Hospitals with the highest numbers of unique participants at HEN events achieved the HEN goals on 19% more eligible topics than

hospitals with the lowest numbers of unique participants at HEN events. Moreover, hospitals sponsoring a Champion Fellow met CMS goals on 13% more topics than hospitals not involved in the Fellowship program.

Participation in HEN Events



Participation In Improvement Leader Fellowship

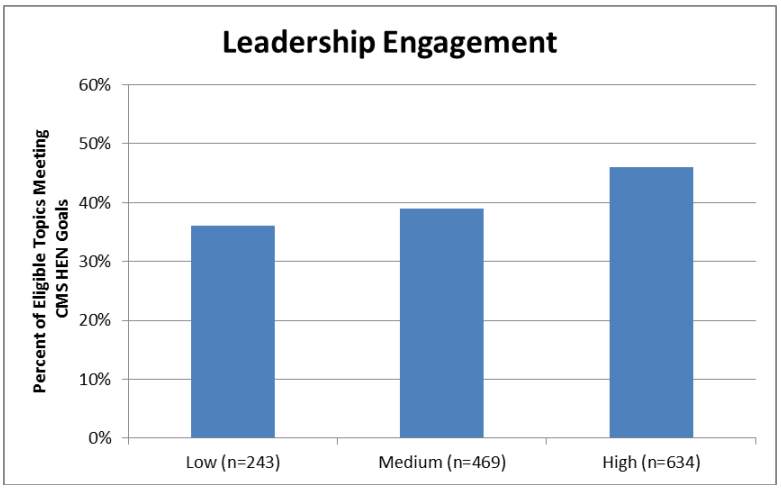


Hospital Implementation Measures

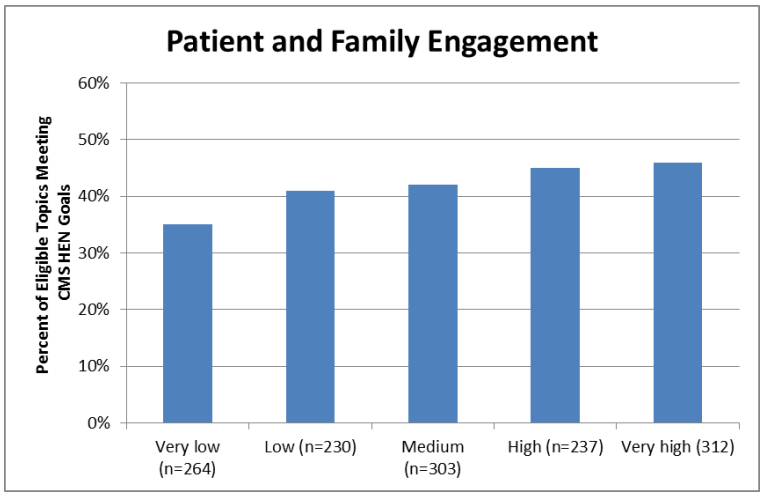
The overall model testing the relationship between the three hospital implementation measures and our primary outcome was significant at $p<.001$ and accounted for 6% of the variance. Leadership engagement and the use of storyboarding were significant individual predicators at $p<.001$ while patient and family engagement was a marginally significant predictor ($p=.053$). Hospitals with the highest levels of leadership engagement met the target goals on 10% more of the eligible topics

than hospitals with the lowest levels of leadership engagement. Hospitals with high or very high scores on the patient engagement level had outcome scores 10% higher than hospitals with low patient engagement scores. A similar pattern was observed for storyboarding; hospitals that have implemented and reported storyboards to track patient harms in their hospital had 15% higher composite measure scores than hospitals that have not implemented this storyboarding.

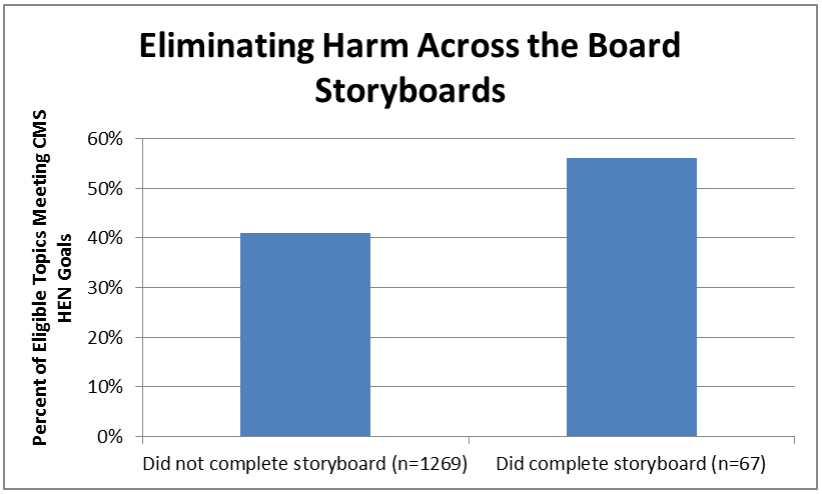
Leadership Engagement



Patient and Family Engagement



Eliminating Harm Across the Board Storyboards

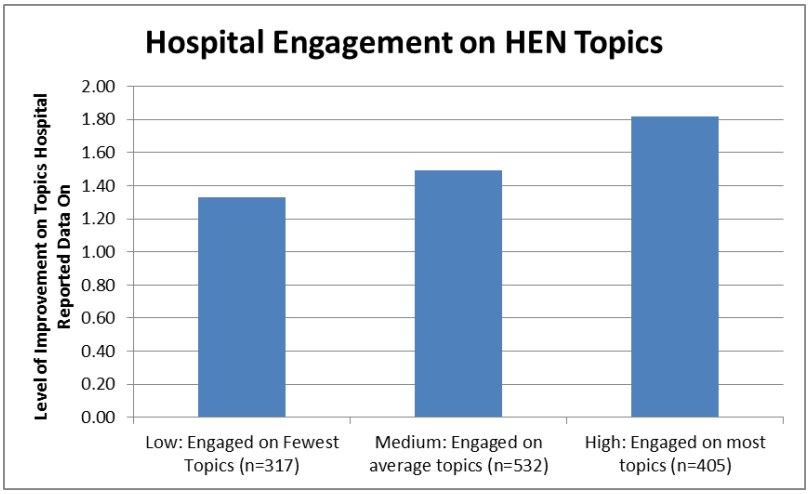


Breadth of HEN Involvement Measures

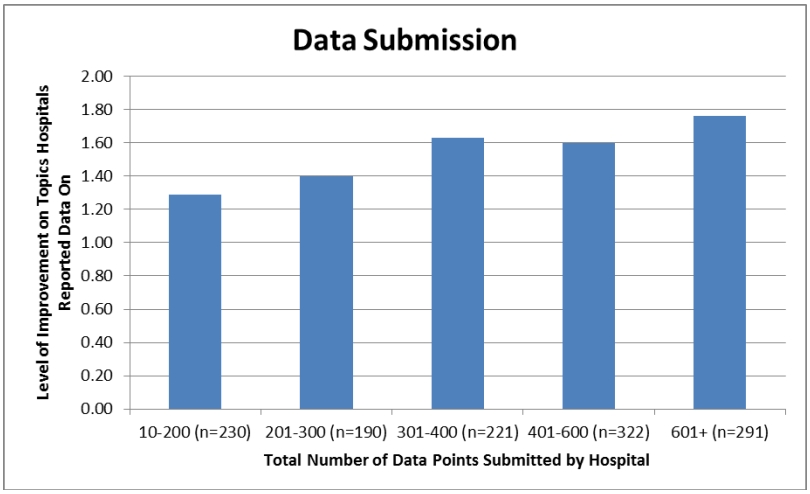
A third approach to assessing the relationship between engagement and positive HEN outcomes involves examining the number of topics hospitals chose to work on and how much data they chose to submit. For these analyses, we could not use the primary outcome measure because hospitals working on fewer topics and submitting less data would inevitably meet a lower percentage of the targeted goals on the CMS composite measure. For these analyses, we assessed the relationship between number of topics and amount of data submitted and a measure of how successful hospitals were on the topics for which they were collecting and reporting data.

The model that included both the number of topics hospitals were working on and the number of data points submitted was highly significant ($p<.001$) and accounted for 21% of the variance in the average amount of improvement observed on measures for which the hospital was reporting data. Both variables were significant independent predictors at $p<.01$. Hospitals working on the highest number of topics were significantly more likely to have met the outcome goals on those topics than hospitals working on the fewest number of topics. Moreover, hospitals reporting more data showed more improvement on targeted topics than hospitals reporting less data.

Number of Topics



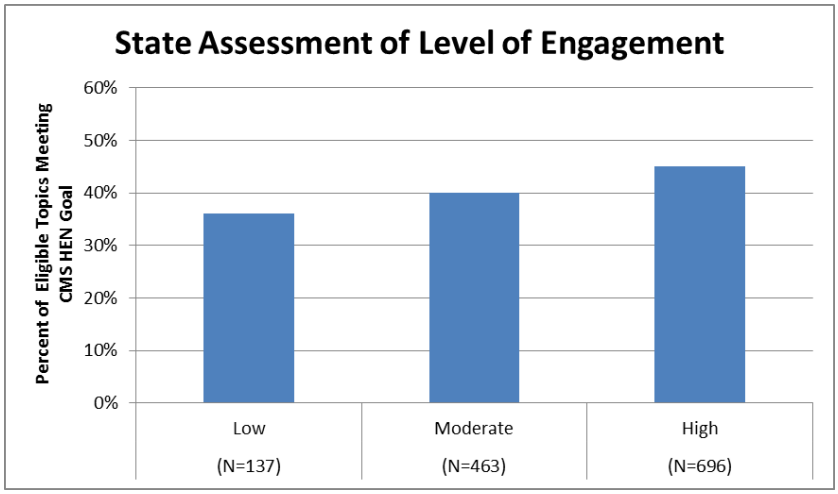
Data Submission



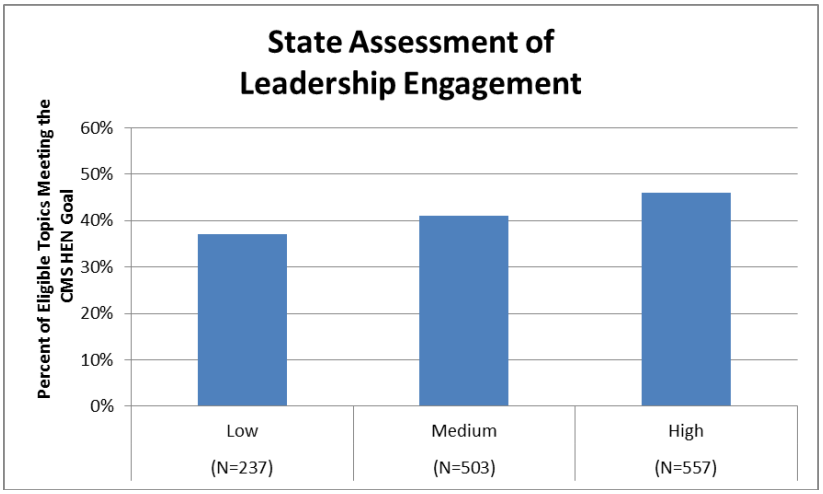
State Hospital Association Assessment of Engagement

SHAs provided their assessments of the extent to which hospitals were fully engaged in HEN, the extent to which hospital leaders were committed to the goals of PfP, and hospitals' reluctance to more fully engage when encouraged to do so by their SHA. Analyses of the model including all three of these predictors was significant at $p<.001$ but only accounted for 3% of the variance in the percent of topics meeting CMS HEN goals. Two measures were significant independent predictors. Hospitals perceived to have the highest levels of overall engagement and leadership engagement had significantly higher outcome scores ($p<.02$). Hospitals perceived as most engaged and with leadership who were most engaged had 7% or higher scores on the outcome measure compared to those with low levels of overall and leadership engagement. Resistance to more extensive engagement was not a significant predictor.

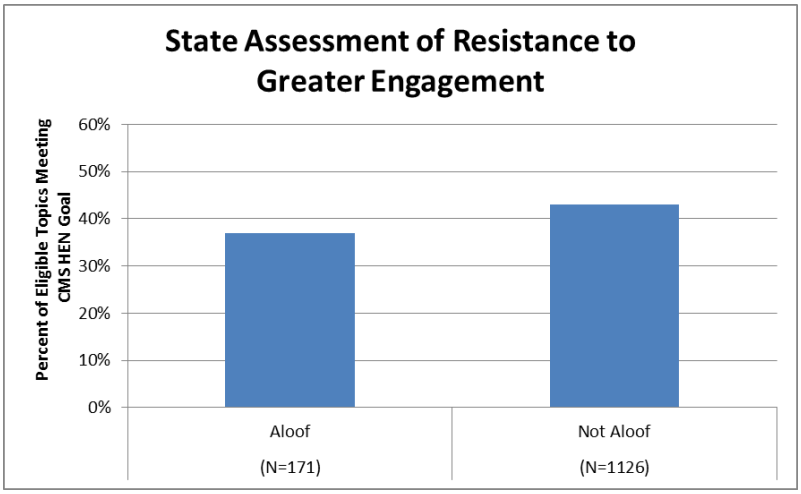
Perceived Overall Hospital Engagement



Perceived Leadership Engagement



Perceived Resistance to Greater Engagement



Conclusions

HRET fully recognizes that these analyses do not provide clear evidence that more extensive engagement in HEN activities caused more improvement in the HEN outcomes. Even though there is a very consistent pattern across each of the engagement measures we used, each of the engagement measures has limitations that may have affected the patterns we observed. Moreover, because it is possible to achieve the goals that the CMS composite outcome measure is based on by either showing significant improvement or by having a very good rate from the outset, it remains possible that the highest initial performers on the topics targeted by HEN were the most engaged. Further analyses should allow us to know whether engagement is linked to improvement or initial high performance.

Even if the set of engagement measures is accepted as valid and further analyses link these measures to improvements on the targeted HEN outcomes, it is still possible that engagement is not causing HEN outcomes to improve. Additional analyses may allow HRET to potentially rule out some alternative explanations and may also provide insights into why some HEN hospitals have progressed more rapidly than others on the important outcomes the PfP is seeking to improve.

Despite these limitations, these analyses represent the first effort to answer a very important question about the relationship between engagement in HEN activities and progress towards achieving the targeted goals of the HEN initiative. The reported results do support the claim that greater engagement in the HEN and HEN activities and priorities is related to higher performance. Across all four engagement domains we examined and all ten of the engagement measures we constructed, higher levels of engagement were associated with higher levels of attainment on the outcome measures targeted by HEN. In a combined analysis using all significant predictors and the outcome measure representing the average amount of improvement observed on topics hospitals were working on, the nine engagement measures accounted for 24% of the variance in this measure. While further analyses can enhance our understanding of this relationship, the results thus far are fully consistent with the hypothesis that more extensive participation in HEN activities leads to better outcomes for patients.