Is the CMS hospital acquired condition reduction program a valid measure of hospital performance?
Authors: Fuller, RL; Goldfield, NI; Averill, RF; Hughes, JS.

Correspondence can be directed to Richard Fuller rfuller@mmm.com

The final version of this article is published in the American Journal of Medical Quality:
Published online before print April 1, 2016, doi: 10.1177/1062860616640883
http://ajm.sagepub.com/content/early/2016/03/30/1062860616640883.abstract

Abstract

In October 2014 the Centers for Medicare and Medicaid Services (CMS) began reducing Medicare payments by one percent for the bottom performing quartile of hospitals under the Hospital-Acquired Condition Reduction Program (HACRP). A tight clustering of HACRP scores around the penalty threshold was observed resulting in 13.2 percent of hospitals being susceptible to a shift in penalty status due to single decile changes in the ranking of any one of the complication or infection measures used in computing the HACRP score. The HACRP score was also found to be significantly correlated with several hospital characteristics including hospital case-mix index. This correlation was not confirmed when an alternative method of measuring hospital complication performance was used. The sensitivity of the HACRP penalties to small changes in performance and correlation of the HACRP score with hospital characteristics call into question the validity of the HACRP measure and method of risk-adjustment.

Background

As required by the Affordable Care Act the Centers for Medicare and Medicaid Services (CMS) in October of 2014 began reducing Medicare payments to hospitals under the Hospital-Acquired Condition Reduction Program (HACRP), a new component of Medicare’s hospital Inpatient Prospective Payment System (IPPS). Under the HACRP, Medicare payments to the 25 percent of hospitals with the highest risk-adjusted rates of selected hospital-acquired complications and infections are uniformly reduced by one percent, for an estimated savings of $330million1.

The HACRP measure is calculated by adding two weighted “domain” scores to form a composite “total” score. The first domain, Patient Safety, uses the Agency for Healthcare Research and Quality (AHRQ) Patient Safety Indicator 90 (PSI-90)2 composite score. The second domain, Healthcare-Associated Infection (HAI), consists of two equally weighted measures provided by the Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN) that compare hospital standardized infection rates for central line associated
bloodstream infections (CLABSI) and catheter associated urinary tract infections (CAUTI) that occurred in intensive care units (ICUs) provided the ICU had at least 20 patients with a central line or had 20 patients with a urinary catheter. All three measures are assigned by decile, with hospitals in the best performing decile (fewest risk-adjusted infections and complications) assigned a score of one and those in the worst decile assigned a score of ten. The overall HAI domain score is calculated as the mean of the CLABSI and CAUTI scores. The final FY2015 HACRP score is a weighted average of the PSI domain (35 percent) and HAI domain (65 percent).

PSI-90 is a weighted composite score across different types of infections and other complications in a single measure. In practice, the weighting of PSI-90 components makes it a measure of surgical complications, pressure ulcers and central line infections. Although central line infections are used in both the PSI and HAI domains, they are calculated differently. In particular, the HAI measure for central line infections (CLABSI) is only applied to ICU patients while PSI-90 includes all patients with a central line infection. Since the HAI CAUTI measure is also ICU specific, hospitals without ICUs have their HACRP score determined solely by their PSI-90. Conversely, hospitals with ICUs reporting CLABSI or CAUTI rates have their HACRP score dominated by the HAI domain (65 percent).

The purpose of this paper is 1) to confirm, as reported in several other studies, that the HACRP score is potentially biased against certain classes of hospitals; 2) to examine whether there are systematic differences in terms of hospital size, characteristics, and patient populations in the score due to the relative contribution of the different domains; 3) to assess the stability of the score and its susceptibility to small changes in complication rates; and 4) to compare the score to another widely used measure of hospital complications in order to determine if any observed bias is due to limitations in the HACRP scoring or due to real hospital performance differences.

Data

HACRP scores were obtained for FY2015 from the CMS website. These are contained within Table 17 of the IPPS FY2015 final rule. Excluded hospitals, those designated N/A under the HACRP heading of “worst performing 25th percentile”, were excluded leaving 3,300 hospitals in the analysis. CMS computed scores for FY2015 using the period January 1 2012 through December 31 2013 for the HAI domain and July 1 2011 through June 30 2013 for the PSI domain.

Component scores within the published 2015 results were retrieved separately from Medicare’s Hospital Compare website. Individual domain scores were available for 3,275 hospitals – 25 fewer than the 3,300 found in Table 17. Provider specific characteristics for indirect medical education, disproportionate share, case-mix index, and hospital bed size were extracted from the IPPS provider impact files for FY2014. The provider file was matched to 3,286 hospitals from Table 17, 14 fewer than the 3,300 assessed by the HACRP. The combined restriction of requiring domain scores and variables from the provider impact file resulted in 3,262 hospitals retained in the analysis.

Method
Contributions of the three domains to total HACRP score were examined for each hospital. Several hospital characteristics were also examined to determine whether they were associated with HACRP scores and the likelihood of penalties:

i. Teaching intensity: measured by the indirect medical education (IME) IPPS payment adjustment factor (resident to bed ratio);

ii. Socioeconomic status: measured by the disproportionate share (DSH) IPPS adjustment (percentage of supplemental security income (SSI)/Medicaid days); and

iii. Hospital case-mix complexity: measured by average MS-DRG relative discharge weight (CMI).

iv. Hospital Size: measured by the average bed size.

The correlation between hospital characteristics and HACRP scores was computed with Pearson’s r. Whether small changes in either the PSI-90 or HAI scores could cause changes in total HACRP score sufficient to cause a change in penalty status was also examined.

To test the robustness of the HACRP performance measure each hospital was reevaluated using an alternative measure employed by the Maryland Health Services Cost Review Commission (HSCRC). The HSCRC has used a more comprehensive measure of potentially preventable complications (PPCs) to adjust payments for hospital-acquired complications under the terms of its Medicare waiver since July 1 2009. The Medicaid programs in Texas and New York adjust hospital payments for complications utilizing the same method of identifying complications and risk-adjustment as Maryland.

In using the Maryland PPC system to examine the correlation between hospital HACRP performance and hospital characteristics, the national average occurrence rate for each PPC was computed to create a PPC norm. On a risk-adjusted basis, the actual number of PPCs in each hospital was compared to the expected number imputed from the norm. For each hospital the difference between the actual and expected number of PPCs was converted to costs using the marginal cost of each PPC, thus establishing the financial impact of excess complications in direct proportion to their associated cost to the hospital. The net financial impact for a hospital was computed by summing the financial impact across each PPC. Good performance (actual number of PPCs less than expected) on some PPCs was allowed to offset poor performance (actual number of PPCs more than expected) on others. Hospital complication performance was measured by the net PPC financial impact. Hospitals with an actual number of PPCs greater than expected received a payment penalty and hospitals with an actual number of PPCs less than expected received a payment bonus. A more detailed description of these methods can be found elsewhere.

Results

Impact of the relative contribution of the different domains to the total HACRP score

Table 1 summarizes the various percent contributions to the HACRP score for hospitals reporting a CAUTI and/or a CLABSI score, or neither, in addition to a PSI-90 score. The variety of permutations results in a variable mix of measures driving the final score used by CMS to compare hospitals. The average hospital bed size and percent of hospitals with a penalty varies considerably depending on whether PSI-90, CLABSI or CAUTI is dominating the computation of the final HACRP score.
Table 1: Distribution of hospitals and domain weighting within the HACRP

<table>
<thead>
<tr>
<th>Domain Weighting</th>
<th>CLABSI</th>
<th>CAUTI</th>
<th>PSI-90</th>
<th>CLABSI</th>
<th>CAUTI</th>
<th>Count Hospital</th>
<th>Percent Hospitals</th>
<th>Average Bed Size</th>
<th>Percent Penalty</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO NO</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>648</td>
<td>19.9%</td>
<td>41.9</td>
<td>13.58%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO YES</td>
<td>35%</td>
<td>0%</td>
<td>65%</td>
<td>353</td>
<td>10.8%</td>
<td>69.9</td>
<td>11.61%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YES NO</td>
<td>35%</td>
<td>65%</td>
<td>0%</td>
<td>7</td>
<td>0.2%</td>
<td>129.7</td>
<td>42.86%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YES YES</td>
<td>35%</td>
<td>32.50%</td>
<td>32.50%</td>
<td>2,254</td>
<td>69.1%</td>
<td>253.7</td>
<td>25.95%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: HACRP scores are obtained from Centers for Medicare & Medicaid Services; Hospital Compare; HAC Reduction Program. 2015. Hospital specific variables are obtained from FY2014 CMS provider impact files.

CLABSI: Central line associated blood stream infection; CAUTI: Catheter-associated urinary tract infection

The consistency of the score is further compromised because when hospital data is insufficient, either because a hospital has no ICU or has less than 20 ICU patients with a central line or a urinary catheter, total scores are reweighted to compensate for the missing measure. Hospitals that incompletely report data are assigned the maximum decile. Hospitals falling into the quartile with the highest scores (worst performing) incur a one percent penalty applied to all IPPS payments.

Table 1 demonstrates that the HACRP score is really not a single uniform score but is composed of several very distinct methods of computing the final score with each of the different methods being applied to hospitals with different characteristics.

A subset of seven hospitals have 65 percent of their HACRP score based solely upon their CLABSI measure. We were reluctant to draw any inference from a group with so few members hence they were excluded from some areas of analysis.

Stability of the HACRP score and its susceptibility to small changes in complication rates

The inherent stability of the HACRP score is questionable because CLABSI and CAUTI are relatively infrequent events (averaging 3.0 and 8.3 per hospital per year, respectively). Indeed, of the hospitals reporting CLABSI and CAUTI events under HACRP, 16.4 percent reported no CLABSI events in a two-year period, while 15.3 percent reported no CAUTI events.

The HACRP penalty quartile begins with scores of 7.025 and above. Hospitals with scores in this range are penalized one percent of their Medicare payments. Hospitals with scores proximate to 7.025 were examined to determine if small changes in PSI-90 or HAIs would cause a change in penalty status.
Figure 1: Distribution of hospitals by HACRP score

Source: CMS FY2015 adjustment file (Table 17). N=3,300.

Hospital HACRP scores tend to concentrate at whole numbers, largely a result of hospitals whose score is determined by the PSI-90 score (Domain 1) only.

Figure 1 summarizes the distribution of hospitals by HACRP score, and demonstrates clustering around whole number scores of 4.0, 5.0, 6.0 and 7.0. This is primarily the result of instances in which hospital scores are calculated using only PSI-90. This applies to 20 percent of hospitals, as reported in Table 1, emphasizing the impact of the different methods of computing the HACRP score. Of particular note is the large number of hospitals (143) lying at the upper bound of quartile 3 (shown in Figure 2) with a score of 7.0 (no penalty) of which 125 (90 percent) have a score based upon PSI-90 alone.

Figure 2 shows the distribution of hospitals by HACRP score within the scoring range of 6.700 to 7.325 where assignment to an alternative decile of a component measure is most likely to move a hospital into or out of penalty. Of the full complement of 3,300 hospitals, 294 (8.9 percent) have scores above 6.700 but below 7.025 (no penalty) and 135 (4.1 percent) have a score between 7.025 and 7.325 (penalty). For these 429 hospitals in proximity to the 7.025 penalty score boundary, a shift of one decile in PSI-90, CLABSI or CAUTI performance in the appropriate direction would change their penalty status. Because CLABSI and CAUTI are
infrequent events, an increase or decrease of a single event can change the hospitals decile performance resulting in a subsequent change in penalty status.

Figure 2: Distribution of hospitals and HACRP scores around the penalty threshold between Quartiles 3 and 4

![Figure 2: Distribution of hospitals and HACRP scores around the penalty threshold between Quartiles 3 and 4](image)

Source: CMS FY2015 adjustment file (Table 17). N=3,300.

Note: Penalty quartile begins at 7.025

**Confirming that the HACRP score is potentially biased against certain classes of hospitals**

Table 2 examines correlation between the characteristics of 3,255 hospitals and their score. The subset of 2,254 hospitals (69 percent) with a total score computed using all three measures have correlation coefficients between the HACRP score and CMI, IME and bed size of $r = .2233$, .2692 and .2455, respectively. The correlation between the HACRP score and DSH is somewhat lower at .1225. A positive correlation, for example that between the score and CMI, means that as CMI increases the HACRP score increases indicating poorer performance and greater likelihood of penalty.
Table 2: Pearson Correlation of HACRP score with hospital CMI, DSH, IME and bed size

<table>
<thead>
<tr>
<th>PSI-90 with</th>
<th>Count Hospital</th>
<th>Case-Mix Index</th>
<th>Disproportionate Share</th>
<th>Indirect Medical Education</th>
<th>Bed Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Pearson r</td>
<td>Mean Pearson r</td>
<td>Mean Pearson r</td>
<td>Mean Pearson r</td>
</tr>
<tr>
<td>Neither CLASBI nor CAUTI</td>
<td>648</td>
<td>1.3681</td>
<td>-0.2358</td>
<td>0.2372</td>
<td>0.1545</td>
</tr>
<tr>
<td>CAUTI only, No CLABSI</td>
<td>353</td>
<td>1.2359</td>
<td>0.0657*</td>
<td>0.2998</td>
<td>-0.0938*</td>
</tr>
<tr>
<td>Both CLABSI and CAUTI</td>
<td>2254</td>
<td>1.5604</td>
<td>0.2233</td>
<td>0.2945</td>
<td>0.1225</td>
</tr>
</tbody>
</table>

* r scores are not statistically significant at the 0.05 level

CLABSI: Central line associated blood stream infection; CAUTI: Catheter-associated urinary tract infection

In contrast, the subset of 648 hospitals with total scores based only upon the PSI-90 tend to be much smaller (average bed size of 41.9), with much lower mean CMI (1.5604 vs. 1.3681) and with observed differences extending to the negative correlation between the HACRP score and CMI (r = -.2358). A negative correlation means that as CMI increases the HACRP score decreases indicating better performance.

It is difficult to interpret the correlation results when the subsets of hospitals are combined because the factors used in computing the HACRP score are different in the subsets of data. The correlations for the 353 hospitals reporting CAUTI only were not statistically significant. Because of these inconsistencies, the remaining analysis focused on the subset of 2,254 hospitals with both PSI-90 and HAI data.

Determining if the observed bias in the HACRP score reflects hospital performance

Table 3 contains the results measuring correlation of the alternative PPC measure with hospital CMI, DSH, IME and bed size for the 2,254 hospitals in which all three measures contributed to their HACRP score.

The correlation between the PPC measure and hospital CMI is not statistically significant, in contrast to the statistically significant HACRP-CMI correlation. The correlation of the net PPC financial impact with IME and bed size drop substantially from .2692 and .2455 for the HACRP score to .1592 and .1075 for PPC. However, correlation of the net PPC financial impact and DSH increases from .1225 for the HACRP score to .1499 for the PPC net financial impact. This could indicate that among larger hospitals there exists a subset that predominately treat poorer
populations and are experiencing higher than expected complication rates as measured by both the HACRP and PPCs.

Table 3: Correlation of HACRP score and PPC actual minus expected with hospital CMI, DSH, IME and bed size in hospitals with HACRP scores comprising all measures: PSI-90; CLABSI & CAUTI.

<table>
<thead>
<tr>
<th>Performance Variable</th>
<th>Hospital Count</th>
<th>Case-Mix Index</th>
<th>Disproportionate Share</th>
<th>Indirect Medical Education</th>
<th>Bed Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>HACRP Score</td>
<td>2254</td>
<td>0.2233</td>
<td>0.1225</td>
<td>0.2692</td>
<td>0.2455</td>
</tr>
<tr>
<td>PPC Net Financial impact</td>
<td>2254</td>
<td>0.0148*</td>
<td>0.1499</td>
<td>0.1592</td>
<td>0.1075</td>
</tr>
</tbody>
</table>

*Pearson r scores are statistically significant at the 0.001 level with the exception of the correlation between case-mix index and PPC net financial impact where p > .05 (P=0.4825)

Discussion

Analysis of the HACRP penalty model raises several serious concerns. The first is its stability in determining which hospitals should be assessed a payment penalty. Very small changes in PSI-90, CLABSI or CAUTI performance can shift the penalty status of a hospital making the HACRP score an unstable indicator of performance. This may be due to the limited scope of the HACRP which is confined to three NQF endorsed measures, the PSI-90, CLABSI and CAUTI. In FY2016 a third measure will be introduced to the HAI domain, Surgical Site Infection: Colon Surgery and Abdominal Hysterectomy (SSI) 15, but the HACRP will continue to be narrowly focused with over-reliance upon ICU measurement.

Secondly, we echo concerns raised by numerous industry stakeholders16–20 in finding HACRP scores to be significantly correlated with CMI, IME and bed size for larger hospitals (Table 2). One interpretation is that large teaching hospitals with a complex patient mix are worse performers in terms of infections and complication. The other interpretation is that the score is not representative of the entire hospital because it is based on a very small number of complications and infections and/or the risk-adjustment used in HACRP is inadequate. The negative correlation between complication performance and hospital CMI was not reconfirmed by the use of PPCs (Table 3) further calling into question the validity and fairness of the HACRP score.

For the 648 (19.9 percent) smaller hospitals measured by PSI-90 alone, the HACRP score is primarily a measure of surgical complications and pressure ulcers. The positive correlation between performance and CMI across these hospitals may mean that lower case-mix complexity is associated with higher rates of complications and thereby a higher PSI-90 score. Alternatively, lower average case-mix complexity may indicate an absence of higher weighted surgical volume within the PSI-90 measure resulting in scores dominated by pressure ulcer performance.

Thirdly, the structure and communication of penalties is central to achieving meaningful behavioral reform. The clustering of HACRP scores that result from summing decile rankings,
the sensitivity of measurement to infrequent events, the subsequent determination of penalties as a product of relative hospital performance only after all other hospital scores are revealed, and the aggregation of disparate measures into a single score, all serve to divorce financial consequences from understanding. For incentives to drive improved performance, investment of money, time and talent needs to be made. As currently structured in the HACRP, the case for this investment, especially if the hospital is not facing any penalty and is either exempted or reports no current adverse events in the limited range of measures, is absent.

High HACRP scores not only result in substantial financial penalties for hospitals but are also intended to serve as signals for patients to choose higher quality hospitals. Under the measure one in five hospitals is assessed under a PSI-90 score in isolation. A similar number of hospitals fall into the “worst” HACRP quartile (score of 7.025 or greater) while having a PSI-90 score of 7.000 or lower. These hospitals, if assessed without the HAI domain, would have avoided a penalty and not been labelled as low quality providers.

Despite its narrow scope, the HACRP penalties are significant and applicable to all Medicare revenue. Our analysis of CMS data reveals that a small performance difference changes whether a hospital receives a HACRP penalty. Hospital-associated infections and complications have been estimated to cost Medicare billions of dollars every year\(^\text{12}\), far in excess of the $330mn dollars targeted to be recouped from the HACRP penalty. For a payment reward or penalty to be fair and/or equitable requires proportionality between performance and penalty. The HACRP delivers large penalties, one percent of payments, to the worst performing quartile of hospitals. That the adjustment makes no distinction between marginal performance differences to assign that penalty is neither proportional nor fair. Moreover, excluding upside incentives and focusing on a narrow range of “worst” performing hospitals means that the best performing hospitals, those that tend to serve as drivers of quality innovation, are almost completely excluded from performance incentives.

**Conclusion**

Fairness and equity in the Medicare payment system should require performance penalties to be designed such that no class of hospital is placed at a disadvantage relative to others due to the services they provide or patients they provide them to. This analysis verifies concerns raised by others that the HACRP score fails to adequately adjust for case-mix and is correlated with certain hospital characteristics. Indeed, the application of a different method of assessing hospital complication performance found no association between performance and hospital CMI. We find HACRP penalties to be sensitive to small changes in infrequent events, poorly structured and offering poor guidance for both patients and hospitals. While this area has the potential to unlock billions of dollars for the hospital industry through averted cost and Medicare through reduced IPPS payments, better and fairer measures need be put in place to stimulate that improvement.
References


