Valuing Improvement in Value-Based Purchasing

William B. Borden, MD; Jan Blustein, MD, PhD

Background—Medicare will soon implement hospital value-based purchasing (VBP) using a scoring system that rewards both achievement (absolute performance) and improvement (performance increase over time). However, improvement is defined so as to give less credit to initial low performers than initial high performers. Because initial low performers are disproportionately hospitals in socioeconomically disadvantaged areas, these institutions stand to lose under Medicare’s VBP proposal.

Methods and Results—We developed an alternative improvement scale and applied it to hospital performance throughout the United States. By using 2005 to 2008 Medicare process measures for acute myocardial infarction (AMI) and heart failure (HF), we calculated hospital scores using Medicare’s proposal and our alternative. Hospital performance scores were compared across 5 locational dimensions of socioeconomic disadvantage: poverty, unemployment, physician shortage, and high school and college graduation rates. Medicare’s proposed scoring system yielded higher overall scores for the most locationally advantaged hospitals for 4 of 5 dimensions in AMI and 2 of 5 dimensions for HF. By using our alternative, differences in overall scores between hospitals in the most and least advantaged areas were attenuated, with locationally advantaged hospitals having higher overall scores for 3 of 5 dimensions for AMI and 1 of 5 dimensions for HF.

Conclusions—Using an alternative VBP formula that reflects the principle of “equal credit for equal improvement” resulted in a more equitable distribution of overall payment scores, which could allow hospitals in both socioeconomically advantaged and disadvantaged areas to succeed under VBP. (Circ Cardiovasc Qual Outcomes. 2012;5:00-00.)

Key Words: pay-for-performance • health policy • Medicare • acute myocardial infarction • heart failure

The push for higher-value health care is driven by the twin stark realities of out-of-control costs and mediocre quality.1 Some see value-based purchasing (VBP) as a promising force to “bend the cost curve,” and this approach will be implemented by the Centers for Medicare & Medicaid Services (CMS) in hospitals nationwide.2–4 In 2007, the CMS delivered a Report to Congress5 that outlined a potential “Performance Assessment Model” method for evaluating and reimbursing hospitals under VBP. In March 2010, the Patient Protection and Affordable Care Act6 was signed into law, instructing the Secretary of Health and Human Services to develop a VBP method by which “the hospital performance score is determined using the higher of its achievement or improvement score for each measure.” The final rule fulfilled this directive by implementing a VBP formula “based on the scoring methodology set forth in the 2007 Report to Congress Performance Assessment Model.”7 Prior work has evaluated the impact of the Performance Assessment Model for process measures8 and forms the basis for ongoing evaluations of the impact of the formula. However, to best understand how the method will affect different hospitals, one must understand the details of the formula itself, which this article will refer to as the “Performance Assessment Model.”

The model (Figure 1A)8 calculates annually for each condition an achievement score (absolute performance) and an improvement score (increase in performance from a prior reporting period). The scale range for the achievement score is the same for every hospital: the lower limit is the 50th percentile of performance for all hospitals, in the previous period; and the upper limit is the mean of the top decile of performance for all hospitals, in the previous period. However, the scale range for the improvement score is unique for each hospital in each year. The lower threshold is that hospital’s performance in the previous period, and the upper limit is the mean of the top decile of performance for all hospitals.
hospitals, in the previous period. For each range, hospitals at or below the lower threshold receive 0 points, those at or above the upper threshold receive 10 points, and those in between receive scaled scores from 1 to 9. The final overall Performance Assessment Model score (blended score) is defined as the greater of the achievement or improvement scores.

Using an Elastic Ruler to Measure Improvement
A corollary of the model is that initial low-performing hospitals have a wider improvement range and, thus, need a greater absolute score increase to achieve the same improvement score as an initial high-performing hospital. In effect, improvement is measured with an “elastic ruler.” For hospitals that have performed poorly in the past, that ruler is maximally stretched, extending downward to their previous period’s performance. For past high performers, the ruler is much shorter.

The Elastic Ruler Hurts Hospitals in Socioeconomically Disadvantaged Locations
The importance of the elastic ruler was underscored in a recent study that applied the Performance Assessment Model to historical hospital performance on the CMS core measures. The study found that hospitals in economically disadvantaged geographic areas were disproportionately low initial scorers. However, those disadvantaged hospitals improved considerably over time, with their improvement exceeding that of nondisadvantaged hospitals in absolute terms. Despite crediting improvement, when the VBP scoring method was...
applied, disadvantaged hospitals received considerably lower blended scores, largely because their greater absolute improvement was discounted by the elastic ruler.

This raises concerns about the redistributive impacts of VBP. Similar concerns about VBP have been voiced in the context of racial equity. Several studies have used simulations and examined pilot projects to predict the financial impacts of VBP on both hospitals and providers. Collectively, this work underscores the complexity of the policy choices and shows how different assumptions and formula choices have sizable and varied impacts on who gains, and who loses, under VBP.

An Alternative to the Elastic Ruler
Rewarding hospitals based on a blend of achievement and improvement is fair and defensible. However, it also seems reasonable to define improvement in a way that gives equal credit for equal improvement. In this article, we describe an alternative formulation of a Performance Assessment Model. Although the clinical process-of-care measures in this study are not identical to those projected for use in the fiscal year 2013 hospital VBP program, they are similar, and they are illustrative of the impact of the VBP method. Our modification uses the same improvement scale for all hospitals (a “wooden ruler,” as described in the Methods section). We test the impact of this modification on the scores received by hospitals in areas of locational advantage and disadvantage.

WHAT IS KNOWN

- Medicare will soon implement a hospital value-based purchasing (VBP) program based on hospital achievement (absolute performance) and improvement (performance increase over time).
- Prior research has shown that undervaluing improvement in the formula has the potential to disproportionately penalize hospitals in socioeconomically disadvantaged areas.

WHAT THE STUDY ADDS

- A modified improvement scoring formula can reflect the principle of “equal credit for equal improvement.”
- This alternative scoring system results in a more equitable distribution of overall payment scores in a VBP simulation.
- VBP scoring that equally rewards both improvement and achievement can allow hospitals in both socioeconomically advantaged and disadvantaged areas to succeed.

Methods
A Modified Performance Assessment Model
As shown in Figure 1B, the modification builds on the original Performance Assessment Model. It uses the same method for assessing achievement with a common scale for all hospitals and brings the improvement calculation into congruence with the achievement calculation. Specifically, the scale range for the current improvement score is defined with the lower limit being the 50th percentile of improvement for all hospitals in the sample from 2005 to 2007 and the upper limit being defined as the mean of the top decile of improvement for all hospitals from 2005 to 2007.

Evaluating the Modified Performance Assessment Model
Study Data Set
To assess the impact of using the Modified Performance Assessment Model, we studied 2 of the conditions to be evaluated in VBP: acute myocardial infarction (AMI) and heart failure (HF). Our data were merged from 3 sources: Medicare Hospital Quality Alliance (HQA) process-of-care data; hospital characteristics and finances from Medicare Cost Reports; and county-level locational data from the Health Resources and Services Administration’s Area Resource File. Included hospitals were those that participated in the Medicare voluntary reporting on HQA measures from 2005 to 2008, with complete reporting of the HQA data for the AMI and HF performance measures during the study period, and for whom the locational characteristics were available.

The HQA process-of-care data, which are publicly reported on the CMS Hospital Compare web site, were used to generate a composite score for AMI and HF. Detailed standards for these measures are published elsewhere and, consistent with a previous study, we selected the following individual measures in developing composite scores: AMI (aspirin on admission, aspirin at discharge, angiotensin-converting enzyme inhibitor for left ventricular dysfunction, β-blocker on admission, and β-blocker at discharge); and HF (assessment of left ventricular function and angiotensin-converting enzyme inhibitor for left ventricular dysfunction). By using a standard method, we generated a single weighted average “composite” score for each condition for each hospital, from 2005 to 2008 in each year, with scores ranging from 0 to 100. This composite HQA score is the input for the Performance Assessment Model scoring.

Prior work details the identification of hospitals that are in socioeconomically “advantaged” and “disadvantaged” counties. Briefly, each hospital is characterized on 5 county-level dimensions: (1) chronicity of local poverty using a modified version of the US Department of Agriculture’s Economic Research Service metric; (2) local unemployment rate according to the 2000 census; (3) designation as a Health Professional Shortage Area (HPSA) for primary care; (4) low education, as designated by the Economic Research Service, based on percentage of individuals with a high school diploma or equivalent; and (5) local prevalence of college graduates, divided into quartiles. Hospitals at the extremes of disadvantage/advantage are “locationally disadvantaged” counties (persistently poor, high unemployment, entire county designated as HPSA, high prevalence of non–high school graduates in the workforce, and lowest quartile college educated) or “locationally advantaged” (never poor, low unemployment, no part of the county designated as HPSA, low prevalence of non–high school graduates in the workforce, and highest quartile college educated).

Statistical Analysis
Individual hospital characteristics and locational dimensions were described. By using the composite performance scores, the absolute score in 2005 and the score improvement from 2005 to 2008 were compared between the least and most advantaged groups, within each locational dimension, using a test of difference in differences by ANOVA, with a Bonferroni correction for multiple comparisons. The distributions of achievement, improvement, and blended scores for 2008 were calculated for both the Original Performance Assessment Model and our Modified Performance Assessment Model, with 2007 data used to define the Achievement Range and 2005 to 2007 data used to define the Improvement Range. Next, mean achievement, improvement, and blended scores for both models were calculated for hospitals by locational dimensions; and bivariate differences in means for blended scores were compared by ANOVA between the least and most advantaged levels, within each locational dimension.
Table. Locational Characteristics of Hospitals and Composite Performance Scores by Locational Dimension in AMI and HF

<table>
<thead>
<tr>
<th>Locational Dimension</th>
<th>Level</th>
<th>AMI Composite Scores by Year, %</th>
<th>HF Composite Scores by Year, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>2205 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronicity of poverty, 1970–2000</td>
<td>Persistently poor</td>
<td>82 (3.7)</td>
<td>87.3</td>
</tr>
<tr>
<td></td>
<td>Intermittently poor</td>
<td>402 (18.2)</td>
<td>89.6</td>
</tr>
<tr>
<td></td>
<td>Never poor</td>
<td>1721 (78.1)</td>
<td>92.9</td>
</tr>
<tr>
<td>Extent of unemployment</td>
<td>High</td>
<td>184 (8.3)</td>
<td>88.9</td>
</tr>
<tr>
<td></td>
<td>Not high</td>
<td>2021 (91.7)</td>
<td>92.4</td>
</tr>
<tr>
<td>Health professional shortage</td>
<td>Entire county</td>
<td>53 (2.4)</td>
<td>86.6</td>
</tr>
<tr>
<td></td>
<td>Portion of county</td>
<td>1477 (67.0)</td>
<td>92.3</td>
</tr>
<tr>
<td></td>
<td>No portion of county</td>
<td>675 (30.6)</td>
<td>92.1</td>
</tr>
<tr>
<td>Non–high school graduates in the workforce</td>
<td>High prevalence</td>
<td>258 (11.7)</td>
<td>88.8</td>
</tr>
<tr>
<td></td>
<td>Not high prevalence</td>
<td>1947 (88.3)</td>
<td>92.5</td>
</tr>
<tr>
<td>College graduates in the workforce, %</td>
<td>&lt;16.6</td>
<td>573 (26.0)</td>
<td>89.0</td>
</tr>
<tr>
<td></td>
<td>16.6–23.2</td>
<td>532 (24.1)</td>
<td>92.3</td>
</tr>
<tr>
<td></td>
<td>23.2–28.3</td>
<td>567 (25.7)</td>
<td>92.7</td>
</tr>
<tr>
<td></td>
<td>≥28.3</td>
<td>533 (24.2)</td>
<td>94.6</td>
</tr>
</tbody>
</table>

AMI indicates acute myocardial infarction; HF, heart failure.

*The improvement in composite scores from 2005 to 2008 was greatest for the least locationally advantaged group, compared with the most advantaged group, within each dimension, with P<0.05 for all comparisons, except for health professional shortage and extent of unemployment for AMI.

†In 2005, the most locationally advantaged group in each locational dimension had higher composite scores, compared with the least locationally advantaged group, with P<0.05 for all comparisons.

Results

The study data set includes 2205 hospitals with complete reporting on HQA measures from 2005 to 2008. Overall, most hospitals were large, with >300 beds (n=1244 [56.4%]), were nonprofit (n=1517 [68.8%]), were located in metropolitan areas (n=1737 [78.8%]), and did not have graduate medical education programs (n=1680 [76.8%] were hospitals with no residency programs). The locational resource levels demonstrate that, for some locational dimensions, there are relatively few hospitals in the most disadvantaged category (Table 1). As noted in a prior study, these “complete reporting” institutions were disproportionately in better-off areas of the nation. Nevertheless, as shown in the Table, there was substantial representation by locationally disadvantaged hospitals in our sample and 22.6% of hospitals in the United States are disadvantaged with respect to at least 1 locational dimension.

The Table also shows the mean scores for the hospitals, stratified by each of the dimensions of disadvantage, for 2005 through 2008, for AMI and HF. For every dimension, the least advantaged hospitals began the period with lower performance scores than the most advantaged hospitals; in all dimensions except HPSA and the extent of unemployment in the AMI measure, the least advantaged hospitals improved more in absolute terms over time than the most advantaged hospitals. The Original Performance Assessment Model yielded statistically significant higher blended scores for the most locationally advantaged hospitals compared with the most locationally disadvantaged hospitals (for all locational dimensions for AMI, except for the extent of HPSA, and for HF in the categories of degree of persistent poverty and quartiles of college graduates; Figure 2A and B). An analysis of the component scores in the Original formula revealed that hospitals in the most advantaged areas did better across all dimensions and both conditions in achievement scores and, although there is less of a disparity, they were similarly favored in scored improvement, reflecting the impact of the “wooden ruler.” Thus, although the most locationally disadvantaged hospitals perform better on improvement in absolute terms, in the Original Performance Assessment Model, their performance is insufficient to match the achievement or the improvement scores of the most locationally advantaged hospitals.

Implementing the Modified Performance Assessment Model resulted in a reduction in differences on blended scores between hospitals in the most and least advantaged areas. Statistically significant differences were no longer evident in AMI blended scores for hospitals in areas with HPSA in the entire county and with a high prevalence of non–high school graduates and for HF blended scores in all dimensions, except for hospitals in areas with HPSA throughout the county (Figure 3A and B). Within the blended score components, the achievement scores were unchanged from the Original, still favoring locationally advantaged hospitals in most dimensions; however, the improvement scores were consistently higher for locationally disadvantaged hospitals for both AMI and HF in all dimensions, reflecting the benefit of the “wooden ruler.” In contrast to the Original Model, then, locationally disadvantaged hospitals fared better in achievement scoring and locationally disadvantaged hospitals achieved greater success in improvement scoring, resulting in a greater parity on the overall blended scores.
Discussion

We modified the Original Performance Assessment Model to reflect the principle of “equal credit for equal improvement.” We then compared performance scores for hospitals using the Original and Modified Models, based on historical performance on the CMS core measures. We showed the impact of this modification by comparing a set of hospitals that are historically low performers (those in socioeconomically disadvantaged regions) against those that are historically high performers (those in advantaged regions). By using the Original Model, hospitals in advantaged regions were winners, with better blended scores. However, using the Modified Model reduced the disparity in blended scores between the 2 groups of hospitals, by giving equal credit for improvement to all hospitals.

Several studies have found that providers serving disadvantaged patients tend to perform worse on process measures of care. This may be because those providers have poorer resources. For example, several studies link provider human resources with performance. A study of pay for performance in the United Kingdom found that low-performing practices disproportionately served income-deprived areas and were disproportionately staffed by older nondomestically trained physicians. Research on hospital nurses in the United States found fewer baccalaureate-prepared registered nurses in areas where the workforce is poorly educated. These same local workforce education deficiencies have been associated with poor hospital performance on the measures reported herein. Other studies have linked provider financial resources and performance. For instance, increased Medicaid revenue dependence and lower total operating margin are both independently associated with diminished performance on the Hospital Compare measures. To be sure, these studies do not prove that diminished resources cause poor performance; also, they do not demonstrate that provider performance would improve, were more resources provided. Furthermore, they do not mean that improvements necessarily require the infusion of new cash. In our data, many of the initially low-performing hospitals improved to a greater degree than their locationally advantaged counterparts, despite existing local workforce and economic disadvantages. This improvement was in the setting of public reporting, suggesting that hospitals respond to incentives beyond the purely financial. Nevertheless, pay for performance rests on the notion that quality improvement requires human and financial resources and that funds are needed to hire and train personnel, organize and support quality improvement teams, and invest in infrastruc-

Figure 2. Original Performance Assessment Model scores by locational dimension. A, Acute myocardial infarction. B, Heart failure.
24,25 It would be surprising if the observed improvements came without financial cost.

To our knowledge, this is the first published study to examine the relationship between alternative definitions of “performance” and the distribution of winners and losers. Much of recent discussion about pay for performance has been about effectiveness and impacts on hospitals of various types. Surprisingly, little attention has been given to the way that improvement is defined.

Caveats
There are several methodological limitations to our analysis. We have used historical data to project likely future trends. The historical data are drawn from the “pay-for-reporting” period, and hospitals may behave differently under the pay-for-performance program. Moreover, the data demonstrate some convergence of scores in recent years between the least and most locationally advantaged hospitals, suggesting that the disparity may be less over time. Any VBP formula must be able to evolve over time, reflecting the dynamic state of quality improvement, and likely would decrease the weighting of improvement relative to achievement as overall performance becomes higher. Other caveats pertain to the identification of hospitals in “locationally disadvantaged areas,” which have been discussed at length elsewhere.8

Policy Implications
Pay for performance offers economic incentives for hospitals to invest in quality improvement by providing a likely return on...
their investment. For example, if a hospital predicts that hiring a quality improvement nurse will drive higher process measure scores and, thus, an incentive bonus greater than the cost of hiring the nurse, then the hospital is likely to make such an investment and achieve better patient care. If, however, a VBP formula rewards initial high performers such that low performers see no opportunity to obtain financial rewards, then low performers will have little financial motivation to invest in quality improvement and may actually decrease services as they prepare for the financial penalties that they may face. Maintaining the fiscal incentive toward quality care for initial high-performing hospitals is also important, and the achievement scoring system provides a mechanism to reward such hospitals. Creating an improvement scale that measures success in a more balanced manner to this achievement scale seems both logical and equitable. As we argue later, this approach may even be more economically efficient.

However, there are several good arguments that would support the use of the elastic ruler. In evaluating the merits of those arguments, it is useful to keep in mind that Medicare’s goal under VBP is to improve the level of care for all beneficiaries. This includes current patients, as well as future patients who might benefit from improvements in quality.

The first argument is that rewarding low performers may institutionalize poor performance. The CMS articulated this rationale in the final rule for VBP, in response to a commenter’s proposal of a fixed improvement scale with a lower “improvement benchmark.” The CMS responded as follows: “we believe establishing a lower [improvement] benchmark would undervalue achievement by lowering the standard by which hospitals may achieve 10 points as well as the importance of improving to the highest level of care. Setting a separate, lower benchmark for the improvement range might also encourage higher achieving hospitals to underperform, as they would be rewarded more highly for achieving a lower level of improvement. A higher benchmark also allows every hospital to improve as much as possible and to the highest level of care.” We believe this means that there is some threshold below which hospitals should not be rewarded, even if they are improving. This viewpoint has merits. Rewarding suboptimal care risks institutionalizing poor performance, and it is hard to strike this balance. Still, it is equally difficult to understand the merits of expecting the highest level of performance from historically underresourced institutions.

A second set of arguments has to do with the relative ease of improvement at the low end of the performance spectrum, and “low hanging fruit.” One variant of this argument says that improved documentation underlies improvement among low performers, rather than improved care. A second variant argues that the marginal costs of delivering better care increase as overall quality increases. We are not aware of any evidence in support of the first variant. However, even if the low hanging fruit is documentation, those improvements require investment, which may, in turn, have positive spillovers for future patients with the same or other conditions. It seems reasonable to provide financial incentives to hospitals that are willing to use their resources to picking this low hanging fruit.

Alternatively, if we believe that the marginal cost of improvement increases with baseline performance, the Medicare program’s most efficient strategy may not be to pay for improve-

ment at the high end. After all, high performers are already eligible for payments under the achievement criterion. Although the marginal costs of improvement increase at the high end, the marginal benefits of improvement remain flat across the spectrum of hospital performance. If a hospital improves a process measure from 70% to 75%, or from 90% to 95%, the absolute improvement is the same and the benefit is the same for those 5% of patients. Thus, the marginal value (defined as marginal benefit over marginal cost) of rewarding improvement is greater for lower-performing hospitals than for higher-performing hospitals. From Medicare’s perspective, although there are competing objectives of supporting market forces and supporting distributive equity, the overarching goal is to achieve the highest national average value. If Medicare can achieve the same marginal benefit at a lower marginal cost, then this is a good deal for Medicare. Such value reasoning supports the use of the “wooden ruler,” favoring improvement for initially lower-performing hospitals.

A third argument takes a strict market approach, arguing that it is best to allow the market to run its course. Future outcomes would be better if (1) low performers were replaced by new institutions or (2) patients from low-performing hospitals sought care from higher-performing hospitals. Although it is beyond our scope to enumerate the market limitations in health care, hospitals differ from commercial firms in several ways. First, for example, there are steep barriers to entry into the hospital market. Hospitals are not coffee shops where if one coffee shop in a neighborhood is serving stale coffee, another coffee shop can quickly open and compete with or replace the existing shop. The financial and regulatory costs to opening a new hospital are substantial, and there is little reason to believe that new institutions could spring up and fill voids, particularly those left by hospitals serving disadvantaged patients. Second, although the current geographic distribution of hospitals may not be optimal, it is unclear that closing low performers would necessarily enhance access. Patients with AMI and HF may not be able to safely travel to a more distant hospital for their care. Finally, the science of performance measurement is new and relatively untested. It seems unwise to peg the survival of institutions on such unproven grounds.

A final argument for the elastic ruler would note that VBP is ill suited as the instrument to solve problems of health care equity. We agree that ensuring resource equity between providers is a complex task and that a multifaceted approach is indicated. Casalino et al. have suggested a variety of other strategies to minimize disparities, including implementing risk adjustment or stratified analyses and assigning rewards specifically aimed at the reduction of disparities. Further steps may involve identifying institutions that particularly need special assistance, such as those with low process measure performance and high mortality and hospital readmission rates. These hospitals can be targeted for focused outreach efforts specific to their individual hospital situations. We think that the first step in this process is to create a level playing field within the VBP formula itself.

Closing Comments

As outlined in the recent health reform legislation, VBP is grounded in the principle that both achievement and improve-
Sources of Funding

Dr Borden receives research funding as a Nanette Laitman Clinical Scholar in Public Health at the Weill Cornell Medical College. Dr Blustein received support under the Wagner School Faculty Research Fund.

Disclosures

None.

References

Valuing Improvement in Value-Based Purchasing
William B. Borden and Jan Blustein

_Circ Cardiovasc Qual Outcomes_. published online February 21, 2012;
_Circulation: Cardiovascular Quality and Outcomes_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/early/2012/02/21/CIRCOUTCOMES.111.962811

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation: Cardiovascular Quality and Outcomes_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation: Cardiovascular Quality and Outcomes_ is online at:
http://circoutcomes.ahajournals.org/subscriptions/