Measuring Outcomes And Efficiency In Medicare Value-Based Purchasing

Medicare’s value-based purchasing plan could be a transformational agent of change.

by Christopher P. Tompkins, Aparna R. Higgins, and Grant A. Ritter

ABSTRACT: The Medicare program may soon adopt value-based purchasing (VBP), in which hospitals could receive incentives that are conditional on meeting specified performance objectives. The authors advocate for a market-oriented framework and direct measures of system-level value that are focused on better outcomes and lower total cost of care. They present a multidimensional framework for measuring outcomes of care and a method to adjust incentive payments based on efficiency. Incremental reforms based on VBP could provoke transformational changes in total patient care by linking payments to value related to the whole patient experience, recognizing shared accountability among providers. [Health Affairs 28, no. 2 (2009): w251–w261 (published online 27 January 2009; 10.1377/hlthaff.28.2.w251)]
operationalize value to the purchaser, which consists of the benefits of services in relation to their cost. Constrained by the availability of measures, the current Medicare hospital VBP model relies on chart-based clinical process-of-care measures, which originally were intended to steer quality improvement (QI) efforts, and self-reported patient experience.\(^1\) The RTC acknowledges the need and intention to expand the measure set to include clinical outcomes and costs.

This paper proposes extensions of the CMS performance assessment model to include clinical outcomes and of the financial incentive model to include cost and thereby promote efficiency and value within the Medicare program. Specifically, we describe a plan for structuring, populating, and scoring the nascent domain of clinical outcome measures and for gauging (taxing) incentive payments in accordance with relative efficiency, to reward value to the purchaser.

**Defining “Value” For Medicare**

One of the central issues in the design of a VBP program is the definition of value. According to the Merriam-Webster online dictionary, *value* is “a fair return or equivalent in goods, services, or money for something exchanged.” In the health care market, value to the purchaser is quality in relation to the cost of care: value can increase by improving quality of services in return for the same payments made to providers, or by lowering the payments made for the same quality of care. The phrase “value-based purchasing” better conveys this second possibility than the former and more common phrase “pay-for-performance” (P4P), which implies additional payments for high or increasing quality.

Value can also increase by improving the quality and efficiency obtained from any given provider, or by shifting patient volume to more efficient providers. Health care purchasers are typically concerned about the total payments made for all covered services delivered to a patient, and they recognize that total payments encompass a range of services delivered over time and throughout the continuum of care. Hence, relative quality, cost, and value are ultimately characteristics of entire delivery systems, involving potentially avoidable and substitutable services.

Of note, the CMS intends for VBP to improve quality of care and remain budget-neutral for Medicare; any additional costs associated with quality improvement must be offset by other reductions in cost. This suggests related objectives—namely, to encourage care coordination and recognition among providers of their shared accountability for patients. The greatest gains in value are likely from prevention, early intervention, and ambulatory management of patients—avoiding the need for emergency department (ED) visits and hospital (re)admissions.

The relative value of hospitals can be measured in terms of clinical quality delivered during the inpatient stay, as well as what happens to patients after discharge. Hospitals are integral components of larger delivery systems, and their total value encompasses their influence over patients and the providers involved before or after discharge. This speaks to the need for a comprehensive approach to Medicare
VBP, extending across different settings and types of providers, while remaining coherent with regard to the functioning of whole delivery systems as well as their constituent parts.

Salient Characteristics Of Outcome Measures

Purchasers care about the health and well-being of patients during and following treatment, and the cost of purchasing such treatment. The clinical choices made by providers (and consumers) are means to those ends; they involve numerous options and reasons for choosing a particular course of action. Improving outcomes may require some providers to choose different clinical processes, or to administer them more consistently. Emphasizing outcomes under VBP would constitute an “open-book test” for providers, who would be motivated to examine their outcomes and their own practices in light of evidence-based or emerging clinical technologies.

With regard to quality, what we propose here is a focus on outcomes. Measures used under VBP should provide as much information as possible to determine the relative quality of individual hospitals. This should help Medicare discern and reward superior performance and help consumers and referring physicians select providers with better outcomes. Examples of such measures are rates of clinical complications, such as secondary infections and other potentially preventable sequelae; premature mortality; and rapid disease progression or acute exacerbations (which can culminate in observable ED visits and readmissions). All of these can be observed and measured after the admission date and during a specified time period after discharge, such as thirty days.

In any one instance or for any one patient, it may not be inferred (solely from administrative data) that a complication was present or that death or readmissions were “to be expected” or preventable. The intent here is to compare groups or cohorts of patients discharged from all relevant hospitals, and to make inferences about relative quality from systematic patterns in which a hospital’s discharged patients have observed rates that differ from case-mix-adjusted expected rates for reference populations of similar patients discharged from other hospitals.

Typically, there are low signal-to-noise ratios associated with individual outcome measures, which means that the information about true quality contained in a measure is mixed with much other variability that cannot be explained easily in statistical models. For example, the acute myocardial infarction (AMI) and heart failure (HF) thirty-day mortality measures currently reported under the RHQDAPU do not discriminate well among hospitals in terms of relative performance. To curtail the effects of noise, the mortality rates were designed to give hospitals the benefit of the doubt and to blend hospital-specific information with average results for other hospitals. A different approach is to combine information from multiple measures, which is integral to the model proposed here.
VBP Outcome Domain Score

To determine financial incentives and for public reporting, the individual outcome measures of a hospital, such as complications and other adverse events, can be combined into scores for patient cohorts (for example, patients with AMI) and for the whole outcomes domain. The resulting aggregate composite scores would reduce multidimensional output to a single dimension for all measured outcomes. This step is essential to generating a Total Performance Score—the input for the “exchange function” in the CMS’s financial incentive model, whose output is intrinsically a single-dimension incentive payment level.

Similarly, public reporting of composite scoring can improve decision making by providing patients and referring physicians with a single, straightforward assessment of hospital quality for a health condition of interest, or aggregated over all health conditions. Meanwhile, the individual constituent measures can be reviewed by hospitals, so that their respective strengths and weaknesses can be understood across a suite of measures and separately by health condition, clinical department, or line of service.

Exhibit 1 provides a framework for calculating scores that can be used for financial incentives and public reporting. The columns of the table represent potential individual measures at a hospital, and the rows represent patient cohorts. The columns will be risk-adjusted rates realized at the hospital for each category of outcome (for example, mortality). The matrix can be calculated using a large number of rows, or types of patients, to refine the statistical models of risk-adjusted ex-

<table>
<thead>
<tr>
<th>Reason for index admission</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Combined outcome score for cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pneumonia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthopedic surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SOURCE: Authors’ analysis.
NOTES: ED is emergency department. AMI is acute myocardial infarction. HF is heart failure.
pected values for each type of outcome (for example, logistic regression models of mortality, Poisson, or negative binomial regressions for ED visits or linear regression for complication rates). Using narrowly defined cohorts would permit more homogeneity of patients; risk adjustment for comorbidities observed in the prior year or salient conditions present upon admission would further improve the accuracy of predicted values for each cohort.

The narrowly defined rows of the matrix can be aggregated to support the public-reporting use of composite scoring. For that purpose, the cohorts must be broad and defined clearly enough to be easily recognized by patients and to allow for sufficient case volume and reliable statistics, yet homogeneous enough that the score reflects the quality of care a cohort patient is likely to receive at the hospital. The exhibit illustrates this with several commonly recognized chronic illness cohorts. A final row is included in the matrix to show overall rates, useful for formulating aggregate hospital-level scores under VBP.

Once calculated, the adjusted rate for each outcome category at each hospital needs to be converted into a number of points to reflect how well the hospital performs with respect to the outcome. This step can use the algorithm described in the CMS’s RTC. The algorithm determines the threshold and benchmark values for each measure based on the distribution across all hospitals, and then awards 0–10 points to each hospital, based on its relative performance.3

For each measure, the number of points earned by a hospital is the greater of its attainment and improvement points. The attainment threshold value is the median of the previous year’s distribution; the benchmark is the mean of the top decile of rates (the average among the best 10 percent of hospitals); and a hospital’s earned points are based proportionally according to where its rate lies between these two values.

By choosing this rule for awarding attainment points, only meritorious performance (above the median) is recognized, and the average number of points provided per hospital is only 2.5 out of 10. This gives hospitals a strong incentive to put forth additional effort on any measure for which they are not yet scoring well. To give positive incentives to all hospitals, regardless of their status quo performance, the VBP plan also calculates and rewards hospitals for significant improvement by comparing their current and recent historical performance. A hospital can earn 0–10 points proportionately according to where its current rate lies between its improvement threshold (its historical value) and the benchmark for this year.

The final step in constructing a composite score, overall or within cohort, is to summarize points across columns, in each case taking the higher of attainment versus improvement points by the hospital.

Constructing composite scores may warrant weighting to account for the severity of the various types of outcomes and to quantify how they should contribute to overall summary values. A hospital with only average performance on minor
complications but far above-average performance on mortality might argue that it should have a (weighted) composite score far higher than that of a hospital with the reverse pattern of rates. Or consider two hospitals with equal overall complication rates of 10 percent, but one hospital’s “minor complication” rate was 9 percent and its “major complication” rate was 1 percent, and the other hospital had the reverse. If columns were weighted equally, and if the columns were mutually exclusive or uncorrelated, both hospitals could easily receive similar composite scores even though they have very different outcomes. Weighting the columns—for example, with increasing integer values as one moves to the right—would seem to be a more reasonable approach.

Even an apparently simplistic choice of equal weights for each column has the appealing property that the composite score will be more influenced by the more severe clinical events, with more robust implications for quality. This points to the interrelationship among the outcome measures. Higher incidence rates of complications during and (manifesting) after the inpatient stay also could trigger ED visits, readmission rates, mortality, and even apparent physician follow-up visits. The greater severity and importance of complications will result in worse scores for a hospital not only in the complication measure but also in other adverse events. Accordingly, more-severe factors will manifest a higher natural weighting in the composite score by affecting single scores and multiple scores.

Weighting columns with assigned values, however, presents its own problems, because there is no obvious way to determine the relative value/harm to assign to the different outcome types (for example, minor complication, major complication, and mortality). The goal here is to define columns in the matrix in such a way that simple addition of a hospital’s points across columns will lead to a reasonable composite score. In Exhibit 1, using the complication rates (columns 2 and 3) as reference points, moving right or left from these two groups, the matrix creates subgroups with either a more favorable outcome (ambulatory physician visits prior to adverse events, in column 1) or with increasingly serious outcomes (ED visits, readmissions, and mortality in columns 4–6, respectively). This could provide a more desirable aggregate score because worse clinical outcomes (major complications) are more likely to be observed in multiple measures. Similarly, ED visits that are systematically related to the index admission may lead to higher probabilities of readmission and death.

Moreover, favorable care coordination will likely be reflected in higher ambulatory follow-up visit rates, fewer ED visits, and fewer hospital readmissions. If we again consider the two hospitals with equal overall complication rates of 10 percent, we would find both hospitals receiving the same points for their “minor complications” rate (column 3), but the hospital with the much lower “major complications” rate (column 4) of 1 percent would likely receive more points than the hospital with the higher rate of 9 percent. Likewise, a hospital with more patients admitted to the ED with a (likely major) complication (column 5) will have
these patients also included in columns 3 and 4, and the hospital’s points in all three columns will be affected. This approach allows the rates for more serious outcomes to affect points awarded across many columns (for example, columns 3, 4, and 5) instead of trying to inflate the points from a single column (column 5) by weighting. Although either normative weights or this alternative natural weighting approach could be used, the latter may be simpler to justify.

**Measuring Cost For Medicare VBP**

For the purposes of VBP, the relevant definition of hospital costs is the price paid for patient care and not the cost of production, because the former represents the actual expense to the taxpayer and would allow policymakers to examine tradeoffs between quality and cost. Given that Medicare pays for services according to price formulas, total costs to Medicare are more directly functions of the volume and intensity of services delivered.

Next, there is the question of which costs should be included in measuring the value of services for hospitalized cohorts. One basic although limited approach would include only Medicare payments associated with the inpatient stay—that is, from admission to discharge. A second approach would expand the scope to include Medicare payments during the thirty-day period postdischarge, irrespective of care setting. Variation in costs during the inpatient stay would be attributable to differences in discretionary procedures, coding of diagnosis-related groups (DRGs), outlier payments, or billings for professional services.

The expanded approach to include utilization postdischarge is more appropriate for VBP for the following reason. To ensure valid comparisons of hospitals on quality-cost metrics, and to allow for quality-cost trade-offs, the underlying units of analysis need to be defined consistently for quality and cost. In the previous section, quality measurement included the thirty-day window after discharge. In addition, this thirty-day window would allow policymakers to further emphasize transitions of care in the postacute period. Care transitions, including rehabilitation, skilled nursing care, and home health services, are frequent in the Medicare population, and specific attention is needed for these hand-offs.4

Expanding the locus of the measurement beyond the hospital can enhance efforts to improve care coordination between the inpatient and other care settings beyond, for example, formulating and transmitting discharge instructions. Acknowledging the extra cost to Medicare associated with avoidable ED visits and readmissions also amplifies the importance to hospitals (and community-based providers) of failing to prevent complications and other sequelae.

Similar to the methodology for measuring quality described above, costs will need to be measured for each of the cohorts for whom quality is measured, as shown previously in Exhibit 1. The costs associated with each cohort will include the costs associated with the index admission and, beginning with the admission date, the facility costs and professional fees incurred during the thirty-day period
postdischarge. These costs will also need to be risk-adjusted to account for patient factors that are beyond the control of the hospital; higher costs would be expected for patients with more complicated illnesses. To obtain a total cost associated with each hospital, the costs across all cohorts will need to be aggregated, similar to the quality performance scores.

**Combining Quality With Cost**

To reward hospitals on differential value—quality combined with cost—one would first determine incentive payments earned for demonstrating levels of quality, and then reduce that amount to penalize high-cost hospitals—in effect, imposing a tax on the earned incentive payments. The use of a tax rate, therefore, helps define a net incentive payment for each hospital predicated on the relative value a hospital manifests to the purchaser.

Here again, more than one option is available to implement this incentive-tax model. One option would be to establish incentive payments based on the total quality performance score for each hospital and the exchange function, and then tax the hospitals separately according to their relative cost. This approach might be inappropriate because it includes potentially invalid comparisons of hospitals that are dissimilar in their quality performance and therefore does not allow Medicare to accurately define, measure, or reward efficiency.

The extended-model approach described here seeks to reward efficiency and incorporates a penalty or tax based on the costs incurred by each hospital relative to a hospital’s total quality score. This approach is an application of the Pigouvian theory of subsidies and taxes to address externality issues, which has widespread application in welfare economics. In this approach, hospitals that incur higher costs for the Medicare program, while providing the same or lower quality of care compared with other hospitals, are penalized. These high-cost hospitals that do not perform better on quality essentially exhibit a kind of free-rider behavior by drawing upon the Trust Fund and, ultimately, the taxpayer.

The following example illustrates the VBP performance assessment model extended to incorporate cost and quality. The numbers shown in Exhibit 2 are for illustration purposes only.

As a first step, each hospital’s total quality score would be computed and an incentive payment calculated based on the total quality score, as described in the RTC and briefly summarized above. The next step would be to group hospitals into cohorts based on having similar or equal total quality scores. For example, hospitals could be categorized according to ten-point increments in their total performance scores (which range from zero to 100 percent). In Exhibit 2, there are three such cohorts. Hospitals A, B, and C are in one cohort (in the range of 81–90), hospital D is in another cohort (71–80), and hospitals E and F are in a third cohort (61–70).

Next, a benchmark cost is defined for each cohort (Exhibit 2). Although it is in
its own cohort, as above, hospital D has been assigned the same benchmark cost as hospitals A, B, and C because it provides lower quality and has a higher cost than hospitals A, B, and C. In other words, the benchmark cost for any cohort cannot be higher than the cost associated with a cohort that provides higher quality. Hospitals would then be taxed at a rate chosen by Medicare based on the difference between a given hospital’s cost and the benchmark cost.

In the example above, the net incentive payment is shown for each hospital. Hospitals A and E receive incentive payments; that received by hospital A is higher than that received by hospital E because of higher quality. Hospital D receives the lowest incentive payment because of its poor quality and high cost.

The tax rate and the incentive payment are policy parameters that can be set by Medicare. A key constraint for the Medicare VBP program is budget-neutrality; this model would allow Medicare to ensure budget-neutrality through the use of two parameters: the rate of incentive payment and the tax rate. More generally, by superimposing cost profiles over hospital cohorts with similar quality scores, the purchaser can base incentive payments on a full determination of value.

**VBP: Change Agent Or Passing Fad?**

Among many private health plans and Medicaid programs, P4P programs have been implemented, in which some payments to providers are made conditional on meeting specified quality standards. Most of these programs have been fairly tepid, which might not be surprising, given the fairly equivalent market clout held by purchasers and providers in this post–managed care era. However, many purchasers are eager to adopt Medicare VBP methods that may be appropriate and promising, to align and compound the incentive structure for providers to improve quality and efficiency.

This paper has examined some measurement issues for VBP that might increase its chances of success—namely, implementing measures that emphasize outcomes

---

**EXHIBIT 2**

Value-Based Purchasing (VBP) Net Incentive Payment Model

<table>
<thead>
<tr>
<th>Hospital</th>
<th>No. of patients</th>
<th>Cost per patient ($)</th>
<th>Total cost ($ thousands)</th>
<th>Total quality score (%)</th>
<th>Benchmark cost ($)</th>
<th>Difference between hospital cost and benchmark cost ($)</th>
<th>Net incentive payment per patient ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>200</td>
<td>300</td>
<td>60</td>
<td>87</td>
<td>300</td>
<td>0</td>
<td>191</td>
</tr>
<tr>
<td>B</td>
<td>300</td>
<td>400</td>
<td>120</td>
<td>87</td>
<td>300</td>
<td>-100</td>
<td>91</td>
</tr>
<tr>
<td>C</td>
<td>500</td>
<td>500</td>
<td>250</td>
<td>87</td>
<td>300</td>
<td>-200</td>
<td>-9</td>
</tr>
<tr>
<td>D</td>
<td>500</td>
<td>600</td>
<td>300</td>
<td>80</td>
<td>300</td>
<td>-300</td>
<td>-109</td>
</tr>
<tr>
<td>E</td>
<td>400</td>
<td>250</td>
<td>100</td>
<td>65</td>
<td>250</td>
<td>0</td>
<td>56</td>
</tr>
<tr>
<td>F</td>
<td>300</td>
<td>400</td>
<td>120</td>
<td>65</td>
<td>250</td>
<td>-150</td>
<td>-94</td>
</tr>
</tbody>
</table>

**Source:** Authors’ analysis.
and shared accountability among providers, and aligning measures of cost along with quality to permit comparisons of cost-effective treatment patterns and reward value.

The Medicare VBP program awaits congressional authorization and associated legislative parameters. There may be a cost to postponing such authorization and implementation, for expectations and support may wane as this approach, like so many others in the past, falls out of fashion or becomes institutionalized by many smaller purchasers with modest features and effects. Similarly, its potential transformational quality may be lessened if political resistance from stakeholders or lack of boldness from policymakers leads to meager financial implications for large differences across hospitals in quality and efficiency.

But so far, there has likely been an advantage to having the extra time. The CMS has been further testing the VBP design and related policy choices. Furthermore, the impending availability of an outcomes domain and methods for addressing cost and efficiency will help focus VBP on intrinsic outcomes and, it is hoped, communicate a mindset of improving and attaining value among providers of hospital and other services.

The force of VBP as a transformational agent of change will depend on a number of factors. The number of dollars made conditional on quality and efficiency is a key starting point. If Medicare places only a small fraction of dollars at risk, then VBP will be self-limiting. Larger risk would strengthen incentives but also the financial threat to hospitals unable or unwilling to improve and the incentives to game the system in order to look good and bypass the risk without making genuine improvements. As mentioned earlier, shifting volume to better providers is itself a net improvement in value for the program; having low-quality and inefficient providers cede market share to other providers may be a desirable outcome.

The VBP plan gives policymakers a number of levers to use as they adjust and steer VBP toward optimal results, such as the dollars at risk. Also of policy importance are the weights given to each of the measure domains. In the RTC, the CMS illustrated the plan with a 70 percent weight for clinical process measures and a 30 percent weight for patient experience. It is hoped that the outcome domain will become populated with suitable and effective measures and given a large and growing weight. Hospitals could earn (back) the dollars at risk, and even more with meritorious performance. Adding the efficiency gauge to the incentive payments will help orient VBP toward value in a fuller sense of the concept.

Providers would be free to measure various aspects of their internal processes that they believe will lead to improved outcomes, without having to rely solely on a predetermined set of measures required by payers. Allowing providers the freedom to measure what they deem to be relevant for improving outcomes can lead to the implementation of innovative practices that might not have emerged in an en-
vironment that promotes “teaching to the test.”

It is not obvious, or even likely, that the same measures are most appropriate for both quality improvement and VBP; rather, there is a need for different types of measures that serve different purposes, with an associated division of labor in terms of who measures what. We have presented a conceptual approach to expanding VBP to incorporate outcomes and cost. Relevant policy parameters—such as the percentage of payment levels withheld to form a savings pool, characteristics of the exchange function, and tax rates applied to incentive payments to account for relative efficiency—would require additional analysis before policymakers can make the final major decisions.

An earlier version of this paper was presented at the Fifteenth Princeton Conference, “Can Payment and Other Innovations Improve the Quality and Value of Health Care?,” sponsored by the Council on Health Care Economics and Policy, 27–29 May 2008, in Princeton, New Jersey. The authors assisted the Centers for Medicare and Medicaid Services (CMS) with the design of the Medicare hospital value-based purchasing program and with the Report to Congress submitted by the CMS on 21 November 2007; however, the views expressed here are not necessarily those of the CMS.

NOTES
1. The Medicare hospital value-based purchasing (VBP) program was designed in light of empirical properties related to measures that were available from the Reporting Hospital Quality Data for Annual Payment Update (RHQDAPU) program, mainly clinical process measures. A patient-experience domain was developed by integrating results from the eighteen-item Consumer Assessment of Healthcare Providers and Systems (CAHPS) Hospital Survey, which encompasses critical aspects of the hospital experience.
3. Details regarding translating measure to points using this approach can be found in Centers for Medicare and Medicaid Services, Report to Congress: Plan to Implement a Medicare Hospital Value-Based Purchasing Program, 21 November 2007, http://www.cms.hhs.gov/AcuteInpatientPPS/downloads/HospitalVBPPlanRTCFINALSUBMITTED2007.pdf (accessed 4 December 2008). The CMS determined that a hospital performing better than half of all reference hospitals (the median value) earned merit for attainment. Meanwhile, the benchmark would be an attainable (empirically observed) standard of excellence.